

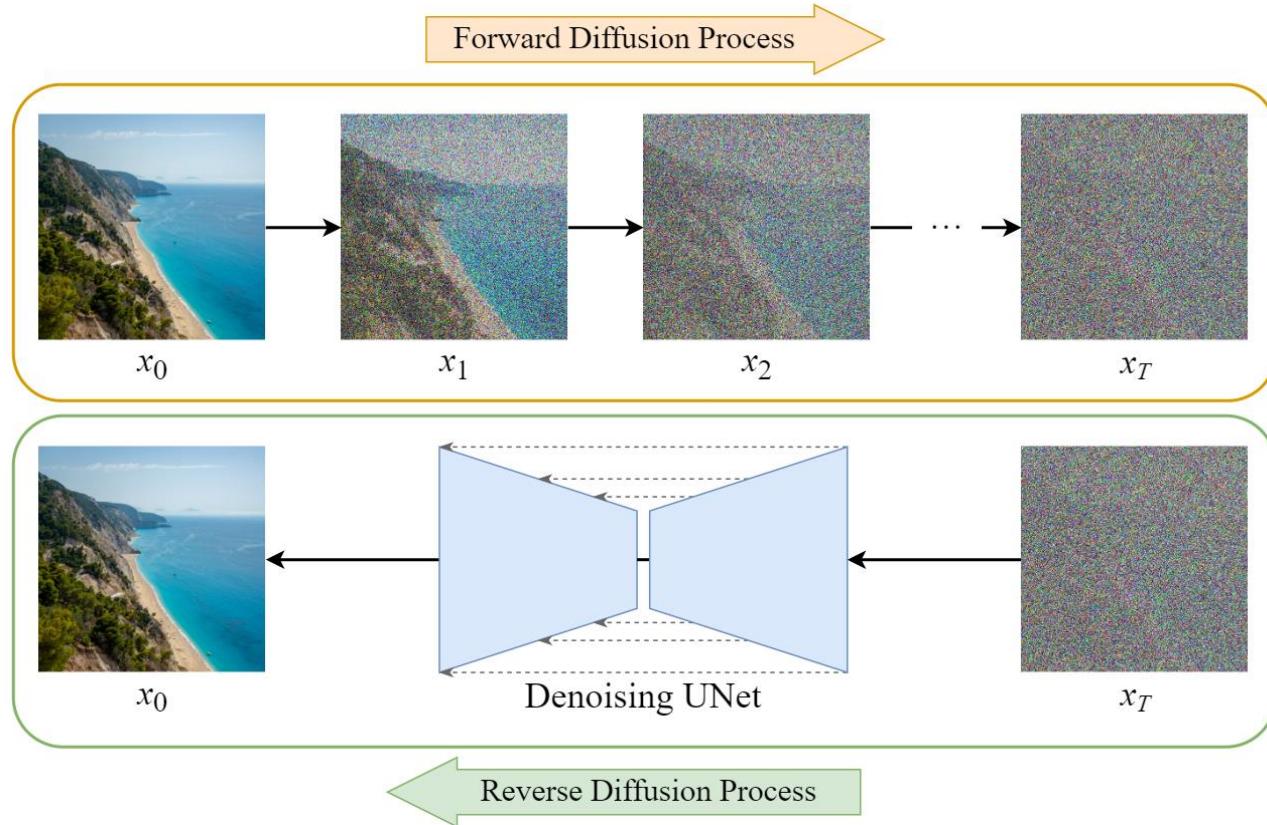
• + ○

IA GENERATIVA PARA HAR

Amparo Diaz

Agenda

- Diffusion Models
- GANS
- LLM



Processo meticoloso de "desdifusão". Primeiro, o modelo adiciona ruído aleatório aos dados reais, criando uma versão corrompida e irreconhecível.

Em seguida, ele treina uma rede neural para reverter esse processo de forma gradual, removendo o ruído e revelando a beleza dos dados originais.

[DALL-E 2](#)

DIFFUSION MODELS

Artigos

- Unsupervised Statistical Feature-Guided Diffusion Model for Sensor-based Human Activity Recognition
- DiffAR: Adaptive Conditional Diffusion Model for Temporal-augmented Human Activity Recognition
- Advancing Diffusion Models for Human Activity Recognition

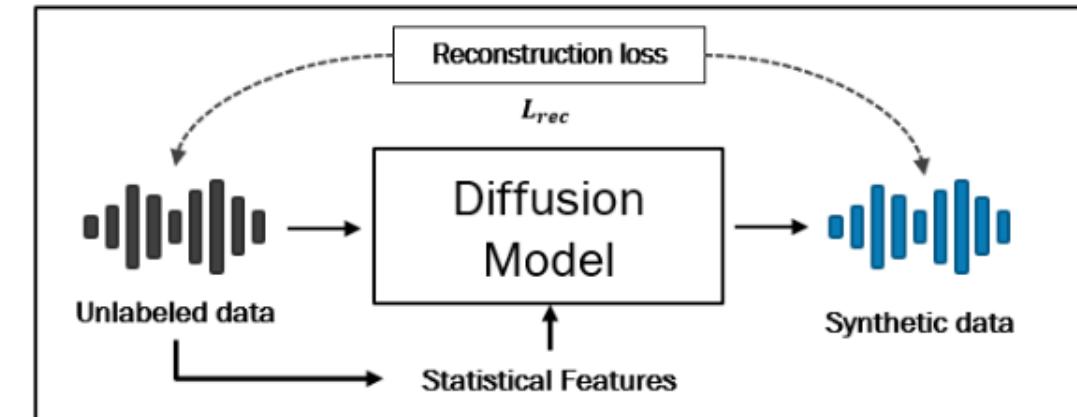
+ Unsupervised Statistical
° Feature-Guided Diffusion
Model for Sensor-based
Human Activity Recognition



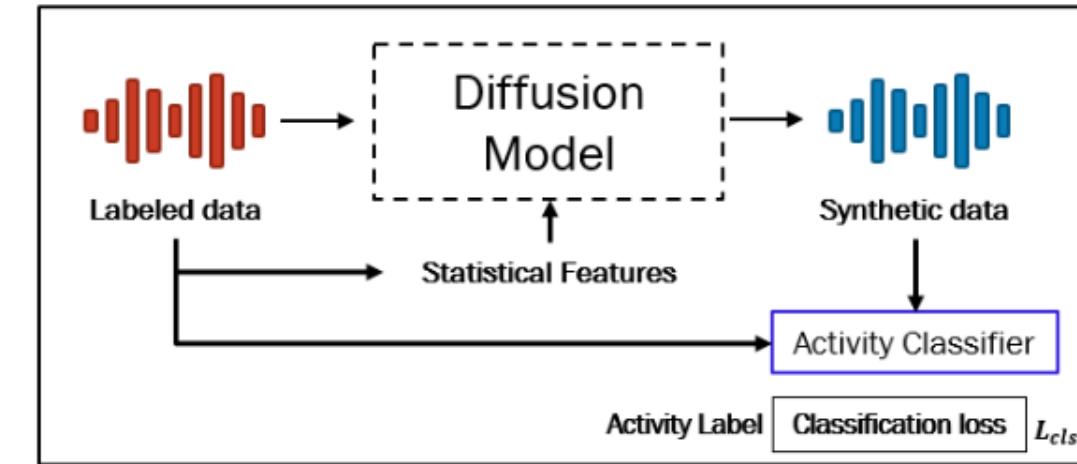
Unsupervised Statistical Feature-Guided Diffusion Model for Sensor-based Human Activity Recognition

- O método proposto consiste em duas etapas principais:
- pré-treinar um modelo de difusão usando dados reais não rotulados. A segunda etapa é pré-treinar um classificador HAR usando dados sintéticos gerados por um modelo de difusão bem treinado (SF-DM)
- posteriormente refinar a classe com dados reais. Para comparação, também propõem um modelo de difusão condicional baseada em classes (CC-DM) que é guiado apenas por informações de rótulo,

[LINK](#)



Step 1. Pretraining Diffusion model on unlabeled data



Step 2. Training activity classifier on generated and real data

Figure 1: Overview of the statistical feature-guided diffusion model (SF-DM).

Unsupervised Statistical Feature-Guided Diffusion Model for Sensor-based Human Activity Recognition

Unsupervised Statistical Feature-Guided Diffusion Model for Sensor-based Human Activity Recognition

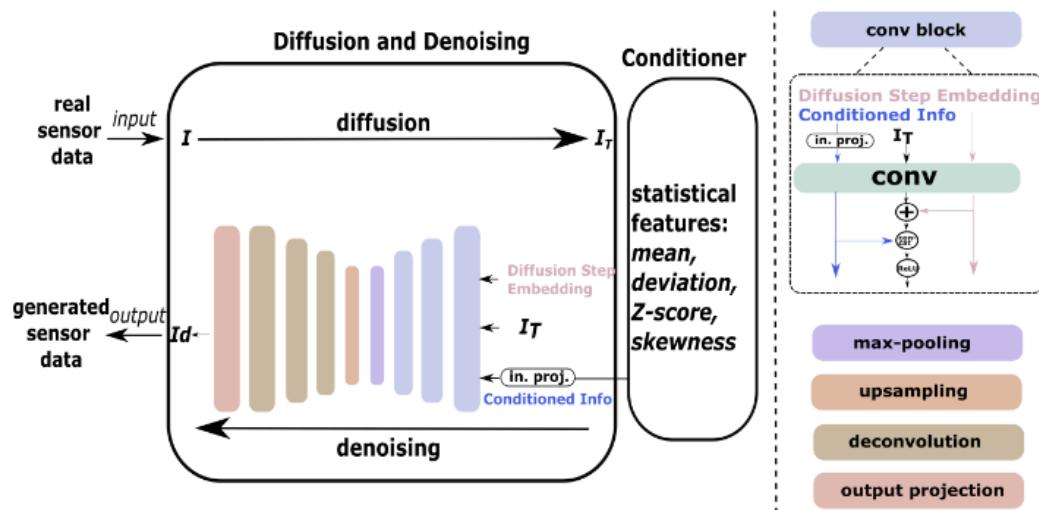


Figure 4: The architecture of the proposed diffusion model.

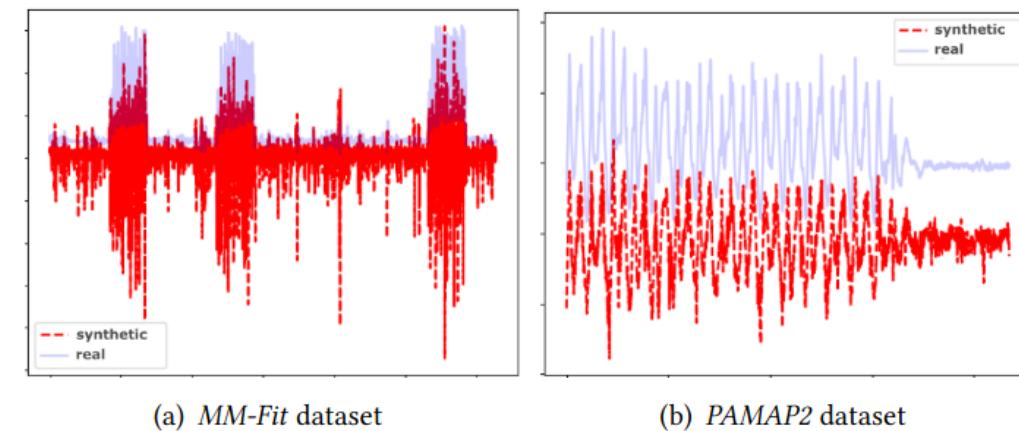
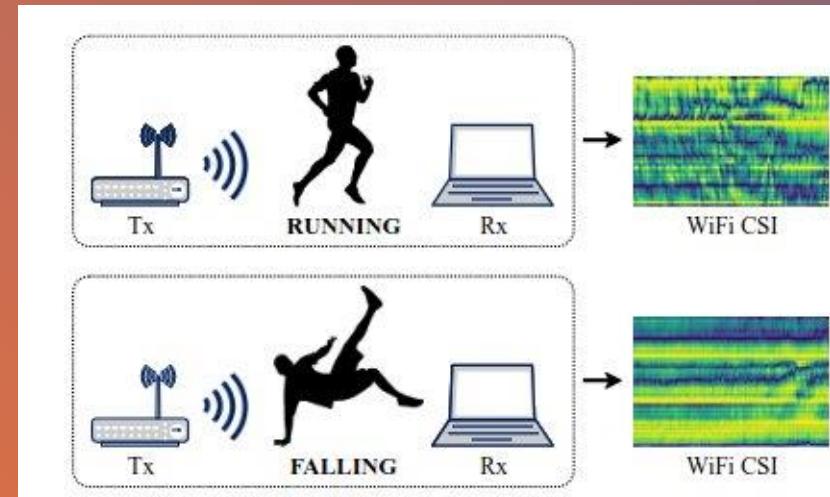


Figure 5: Examples of generated sensor data. The data in red indicates the synthetic sensor data from the diffusion model, while the blue one represents the real sensor data.

DiffAR: Adaptive Conditional Diffusion Model for Temporal-augmented Human Activity Recognition



+

○

DiffAR: Adaptive Conditional Diffusion Model for Temporal-augmented Human Activity Recognition

Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)

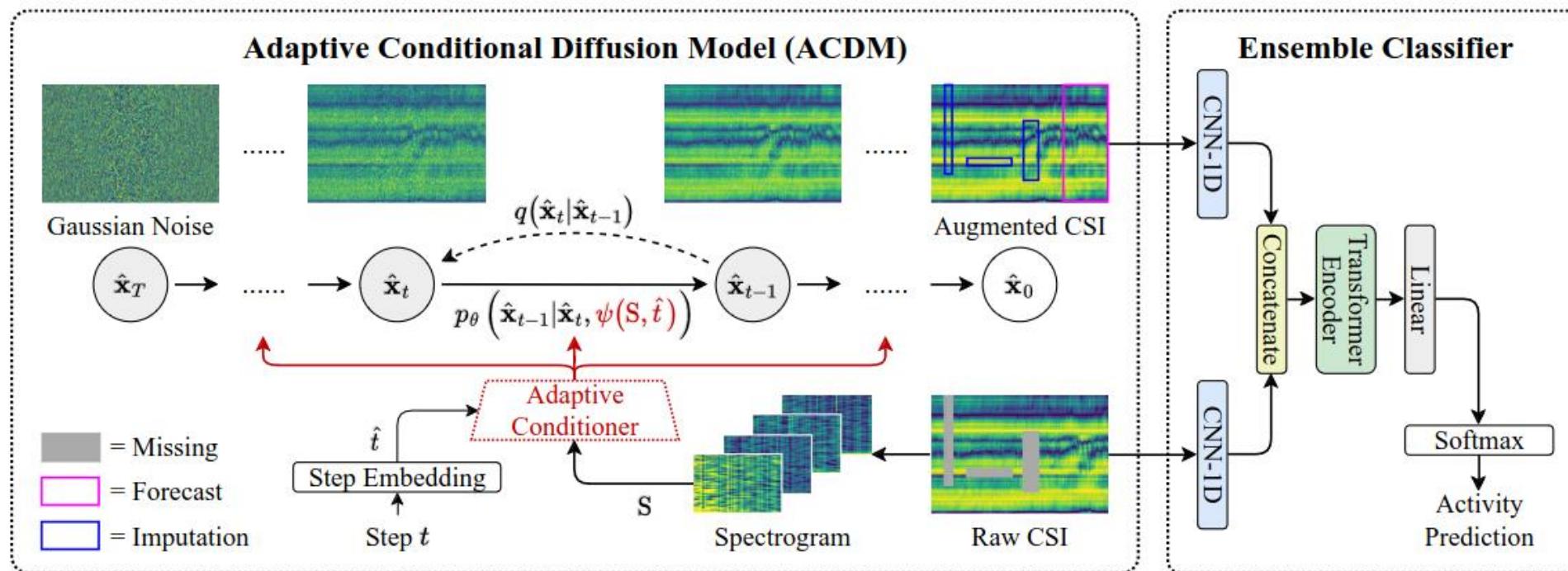


Figure 3: Overview of the proposed DiffAR.

DiffAR: Adaptive Conditional Diffusion Model for Temporal-augmented Human Activity Recognition

Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)

Methods	Ratio		Office Acc.	SignFi		Interactions Acc.	Widar 3.0							
	λ_{fc}	λ_{im}		WP	F1		Acc.	WP	F1					
<i>Baselines</i>														
ST-RF	0.0	0.0	89.95	90.58	85.61	84.06	88.98	82.37	75.42	75.23	74.58	55.42	56.06	55.39
LSTM	0.0	0.0	94.44	94.53	91.50	88.40	90.99	86.15	82.71	82.67	81.89	67.13	67.07	66.73
CNN-1D	0.0	0.0	95.45	95.86	93.47	97.34	98.04	97.36	82.92	83.05	82.45	77.98	78.69	78.19
CNN-2D	0.0	0.0	96.46	96.67	94.85	97.34	97.77	96.19	90.21	90.90	89.86	87.99	88.02	87.95
CNN-LSTM	0.0	0.0	91.92	92.32	88.58	88.04	91.04	87.37	76.25	77.51	75.64	70.52	70.91	70.63
ABLSTM	0.0	0.0	95.96	96.13	94.92	96.38	97.04	95.60	86.46	86.88	85.92	73.47	73.50	73.36
THAT	0.0	0.0	96.97	97.02	95.85	96.74	97.42	96.29	90.63	91.19	90.30	90.04	90.06	90.01
<i>Forecast</i>														
THAT + TimeGrad	0.2	0.0	97.49	97.51	96.25	95.65	96.76	95.58	90.83	91.36	90.68	91.22	91.32	91.21
THAT + DiffWave	0.2	0.0	96.98	97.04	95.99	97.10	97.92	96.28	90.83	91.13	90.75	91.39	91.53	91.40
DiffAR (Ours)	0.2	0.0	97.99	98.10	97.16	98.07	98.78	97.72	94.17	94.37	94.02	91.78	91.76	91.71
<i>Imputation</i>														
CNN-1D + DiffWave	0.0	0.5	95.48	95.46	93.76	97.58	98.01	96.92	85.42	85.62	85.07	81.39	81.65	81.14
THAT + DiffWave	0.0	0.5	96.48	96.40	94.46	96.98	97.75	95.47	90.63	90.78	90.44	91.11	91.09	91.07
DiffAR (Ours)	0.0	0.5	97.99	97.95	96.82	97.95	98.48	97.39	93.75	93.96	93.40	91.67	91.67	91.61
<i>Forecast + Imputation</i>														
CNN-1D + DiffWave	0.2	0.5	96.98	97.01	95.54	97.71	98.03	96.80	85.83	86.18	85.46	82.17	82.41	82.03
THAT + DiffWave	0.2	0.5	96.48	96.76	95.03	97.22	97.48	95.60	90.63	90.97	90.22	90.61	90.62	90.58
DiffAR (Ours)	0.2	0.5	98.49	98.54	98.22	98.19	98.59	98.22	94.58	94.67	94.50	92.06	92.19	92.04

Table 3: The recognition performance (unit: %) of CSI-based HAR methods in terms of Accuracy (Acc.), Weighted Precision (WP) and F1 score. Higher results indicate better performance. **Bold** highlights the best results.

[link](#) [codigo](#)

Advancing Diffusion Models for Human Activity Recognition



Advancing Diffusion Models for Human Activity Recognition

Neste estudo, foi proposto um método para geração de dados de reconhecimento e classificação de atividades humanas (HAR) utilizando um modelo de difusão tabular capaz de lidar tanto com dados contínuos quanto discretos.

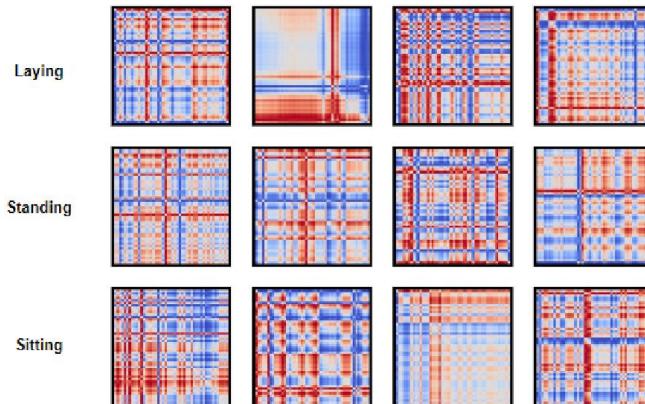
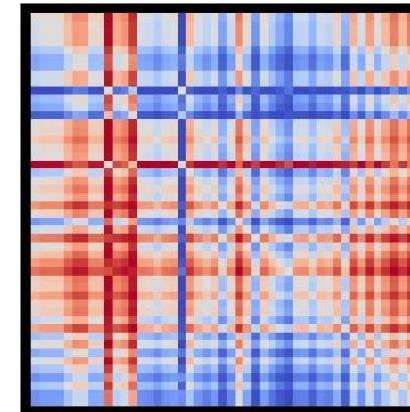
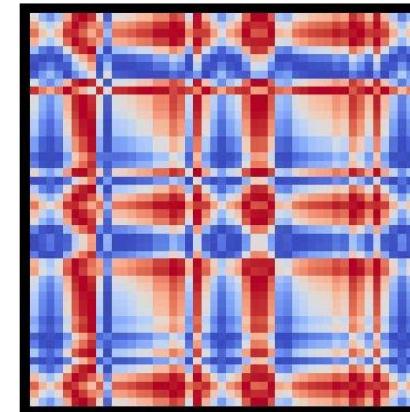


Figure 9: Randomized sample of GAFs generated on Standing, Laying, Sitting HAR X-A data.

Figure 10: Randomized sample of GAFs generated on Walking, Downstairs, Upstairs HAR X-Axis data.



(a) Example of GAF image describing passive action (Sitting, X-Axis, Timestep: 4017).



(b) Example of GAF image describing active action (Upstairs, X-Axis, Timestep: 2146).

Figure 11: Passive vs. Active Action

individual GAFs by axis. We then assign a color channel to each axis (X - red, Y - green, Z - blue) and combine them into one singular RGB-colored GAF to better visualize all three axes in one representation.

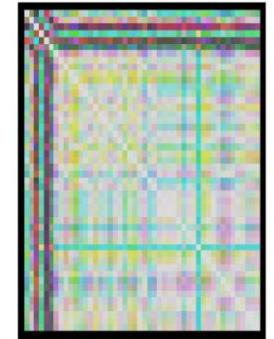


Figure 12: Example of combined axis GAF by RGB color.

[link](#)

Advancing Diffusion Models for Human Activity Recognition

GAFs did not capture the underlying patterns in the data as effectively as the original time series data. Because of these results, we elected to use statistical features as input for our tabular diffusion model.

F1 Scores by Input and Model		
Input Type	Raw Time Series Data	GAFs
Model Type	Random Forest	CNN
F1 Score	0.919 (± 0.007)	0.239 (± 0.023)

Table 1: Results of using raw time series data and GAFs as input to the classifier.

[link](#)

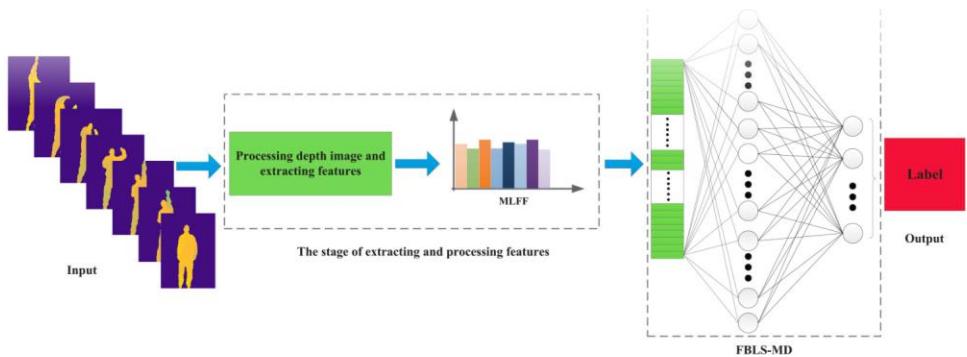


Figure 1. The overview of the proposed method. A new descriptor (i.e. MLFF) is designed for depth images representation. Then FBLS-MD is adopted as the classifier.

Depth-based human activity recognition via multi-level fused features and fast broad learning system

LLM

- **Vantagens da abordagem** de uso de LLM em séries temporais:
 - LLMs podem capturar padrões complexos, como sazonalidade e tendências, que podem ser difíceis de modelar com métodos tradicionais.
 - LLMs podem lidar naturalmente com dados ausentes e informações textuais, que são comuns em séries temporais reais.
 - LLMs podem potencialmente explicar suas previsões, fornecendo insights sobre os fatores que influenciam a série temporal.

Limitações

- Modelos maiores geralmente apresentam melhor desempenho, mas podem ser computacionalmente caros para treinar e usar.
- A calibração de incerteza nas previsões de LLM pode não ser confiável, o que pode limitar sua utilidade em aplicações críticas.

Large Language Models for Time Series: A Survey

Taxonomia de LLMs para Análise de Séries Temporais

Pipeline típico de PLN (Processamento de Linguagem Natural) orientado por LLMs (Large Language Models) com cinco etapas: texto de entrada, tokenização, incorporação, LLM e saída.

Cada categoria da nossa taxonomia visa uma etapa específica desse pipeline:

- **Prompting (Etapa de Entrada):** Trata dados de séries temporais como texto bruto e direciona LLMs diretamente para processá-las.
- **Quantificação de Séries Temporais (Etapa de Tokenização):** Discretiza as séries temporais em tokens especiais para que as LLMs possam processá-las.
- **Alinhamento (Etapa de Incorporação):** Projeta um codificador de séries temporais para alinhar incorporações de séries temporais com o espaço da linguagem.
- **Visão como Ponte (Etapa LLM):** Conecta séries temporais a Modelos de Linguagem e Visão (VLM) empregando representações visuais como ponte.
- **Integração de Ferramentas (Etapa de Saída):** Utiliza modelos de linguagem para gerar ferramentas que auxiliem a análise de séries temporais.

LLM->Prompting

Ideia central: Ao converter dados de séries temporais em formato de texto, os LLMs podem prever valores futuros diretamente, atingindo desempenho comparável a modelos de previsão especializados.

LLM->Prompting

- Evaluating Large Language Models as Virtual Annotators for Time-series Physical Sensing Data

Este artigo propõe uma nova abordagem para anotação de dados de sensores baseada em LLMs.

duas fases:

- Fase 1:** Analisamos os desafios enfrentados por um LLM (como o GPT-4) ao interpretar dados brutos de sensores.
- Fase 2:** Exploramos a possibilidade de codificar dados brutos usando técnicas SSL e utilizar os dados codificados para obter anotações do LLM.

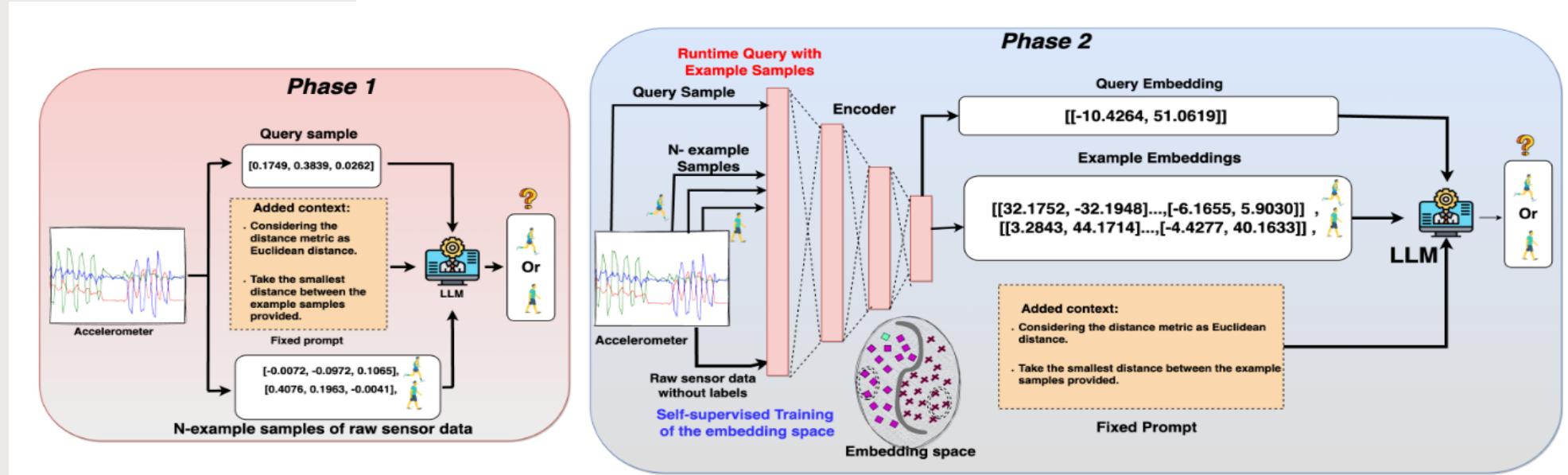


Fig. 6. The broad overview of the setup that the designed study investigates.

LLM->Prompting

- Evaluating Large Language Models as Virtual Annotators for Time-series Physical Sensing Data

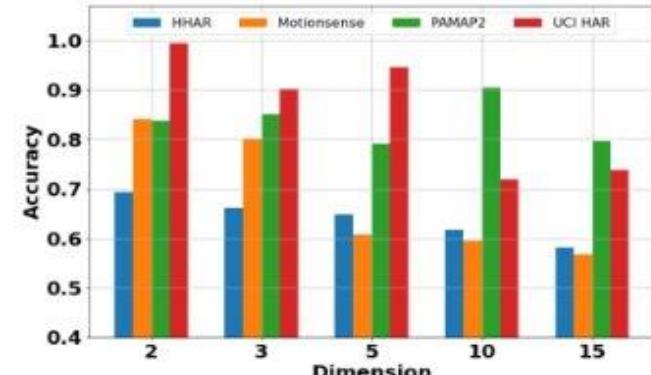
Table 2. Impact of including distance metric as a context information in the query. The responses clearly show how GPT-4 then uses the distance metric as the basis for reasoning regarding the generated response.

Raw Sensor Data	Query	The following given triaxial accelerometer readings correspond to "jogging": [[0.0223 0.1115 0.3571]...[0.161 0.1497 -0.2369]] and the following given triaxial accelerometer readings correspond to "walk upstairs": [[0.2615 0.118 -0.0637]...[-0.0322 0.0437 0.057]]; classify the triaxial accelerometer reading [-0.0933 0.0165 0.0609] as either "jogging" or "walking upstairs" provided that this data is coming from the smartphone stored in the user's trousers' front pocket considering the minimum distance to the example triaxial accelerometer readings provided that the distance metric chosen is euclidean distance. Answer in one word.
	Response	The triaxial accelerometer reading [- 0.0933 , 0.0165 , 0.0609] [-0.0933, 0.0165, 0.0609] is classified as "walk upstairs" based on the minimum Euclidean distance to the example triaxial accelerometer readings provided.
	Reasoning	The classification of the triaxial accelerometer reading [- 0.0933 , 0.0165 , 0.0609] [-0.0933,0.0165,0.0609] as walk upstairs" is based on the principle of the nearest neighbor algorithm, using the Euclidean distance as the metric for similarity. This distance metric was chosen to compare the given accelerometer reading with each of the example readings from both activities ("jogging" and "walking upstairs"). Minimum Distance: For each activity, I calculated the Euclidean distance from the given reading to all example readings of that activity. Then, I found the minimum distance to each activity, determining the closest match for the given reading. Classification: The activity whose examples have the minimum distance to the given reading is considered the most similar or the most likely classification for that reading. In this case, the minimum distance between the given reading and the examples of "walk upstairs" was smaller than that to the examples of "jogging". This indicates that the given reading is more similar to those recorded during the activity of walking upstairs, leading to its classification as "walk upstairs". This method is effective for classifying data points based on similarity to known examples, and it's particularly useful in contexts like activity recognition from accelerometer data, where patterns of motion can be distinctive but vary slightly from person to person.

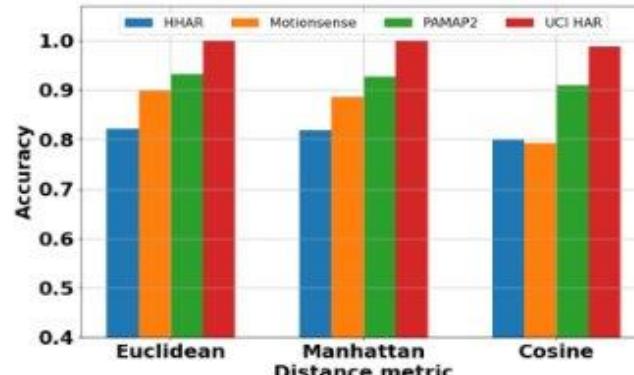
TFC Encoding	Query	The following given embeddings correspond to "walking upstairs": [[4.256844 -10.712115] ... [39.49697 -33.193176]] and the following given embeddings correspond to "jogging": [[38.99644 -7.460116] ... [40.007973 4.696819]]; classify the embedding [2.0370352 3.47346] as either "walking upstairs" or "jogging" considering the minimum distance to the example embeddings provided that the distance metric chosen is euclidean distance. Answer in one word.
	Response	The embedding [2.0370352 , 3.47346] is classified as "walking_upstairs" based on the minimum Euclidean distance to the example embeddings provided for "walking upstairs" and "jogging". This classification indicates that the given embedding is closer to the set of embeddings associated with "walking upstairs" than it is to those associated with "jogging".
	Reasoning	To determine the classification of the embedding [2.0370352 , 3.47346] [2.0370352,3.47346], I calculated the Euclidean distance between this point and each of the given embeddings for "walking upstairs" and "jogging". After computing the distances to all provided embeddings, I identified the minimum distance from the point to the embeddings of each activity category. The logic behind this approach is that an embedding is more likely to belong to the category to which it is closest in terms of distance in the embedding space. This space is structured so that similar activities (in terms of sensor readings, motions, or whatever features these embeddings represent) are closer together. The results showed that the closest distance between the point [2.0370352 , 3.47346] and the "walking upstairs" embeddings was significantly smaller than the closest distance to the "jogging" embeddings. This implies that the point is more similar to, or shares more characteristics with, the "walking upstairs" activity than with "jogging". Therefore, based on the principle of nearest neighbors (where an entity is more likely to be similar to its nearest neighbors in a feature space), the embedding was classified as "walking upstairs". This method is a fundamental concept in machine learning for classifying data points based on the known classification of other points nearby in the feature space.

LLM->Prompting

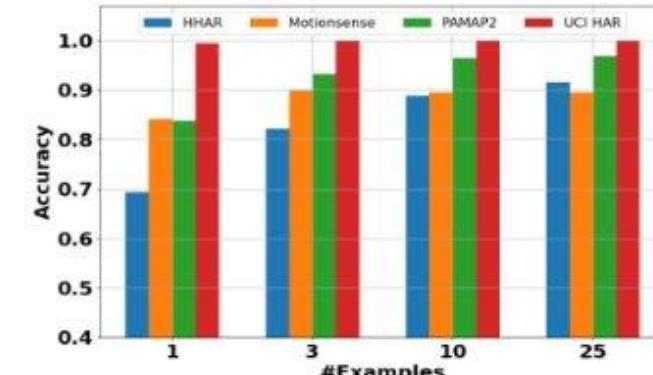
- Evaluating Large Language Models as Virtual Annotators for Time-series Physical Sensing Data



(a)



(b)



(c)

Fig. 11. Accuracy of the responses obtained from GPT-4 using the encoded output from a pre-trained encoder trained using the TFC approach. Here, we show the results by varying – (a) the dimensionality of the embeddings, (b) the distance metric for comparing the embedding with the example embeddings included as a part of the query, and (c) the number of examples.

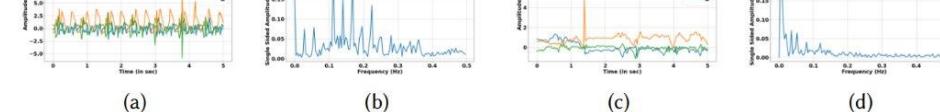


Fig. 7. Time and frequency domain analysis of a time-series data from MotionSense dataset across two different activity classes (a) and (b) correspond to time and frequency domain information for the activity “Jogging” whereas (c) and (d) correspond to time and frequency domain information for the activity “Upstairs”.

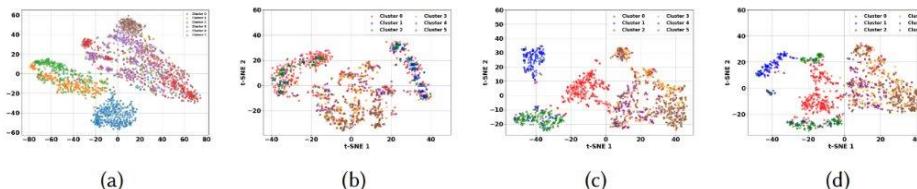


Fig. 8. t-SNE plots of TFC embeddings for the HHAR [2] across – (a) time domain, (b) frequency domain, and (c) both time and frequency domain embeddings concatenated.

LLM->Prompting



Leveraging Language Foundation Models for Human Mobility Forecastin



Date	Paper	Institute	Conference	Domain	Task	LLM
11 Sep 2022	Leveraging Language Foundation Models for Human Mobility Forecasting	University of New South Wales	SIGSPATIAL'22	Mobility	Forecasting	BERT, BoBERTa, GPT-2, etc.

[codigo](#)

Este trabalho propõe um novo método para prever mobilidade humana usando linguagem natural. Ao invés de dados numéricos brutos, o modelo analisa texto descrevendo padrões temporais.

[artigo](#)

Table 2: Pre-trained language models explored in the proposed AuxMobLCast.

	Model	HuggingFace Configuration
Encoder	BERT	https://huggingface.co/bert-base-uncased
	RoBERTa	https://huggingface.co/roberta-base
	XLNet	https://huggingface.co/xlnet-base-cased
Decoder	GPT-2	https://huggingface.co/gpt2

LLM->Prompting

Penetrative AI: Making LLMs Comprehend the Physical World

"IA Penetrativa", um novo conceito relacionado à exploração do uso do conhecimento comum de modelos de linguagem grandes como base para realizar tarefas de percepção e intervenção no mundo real em sistemas ciber-físicos

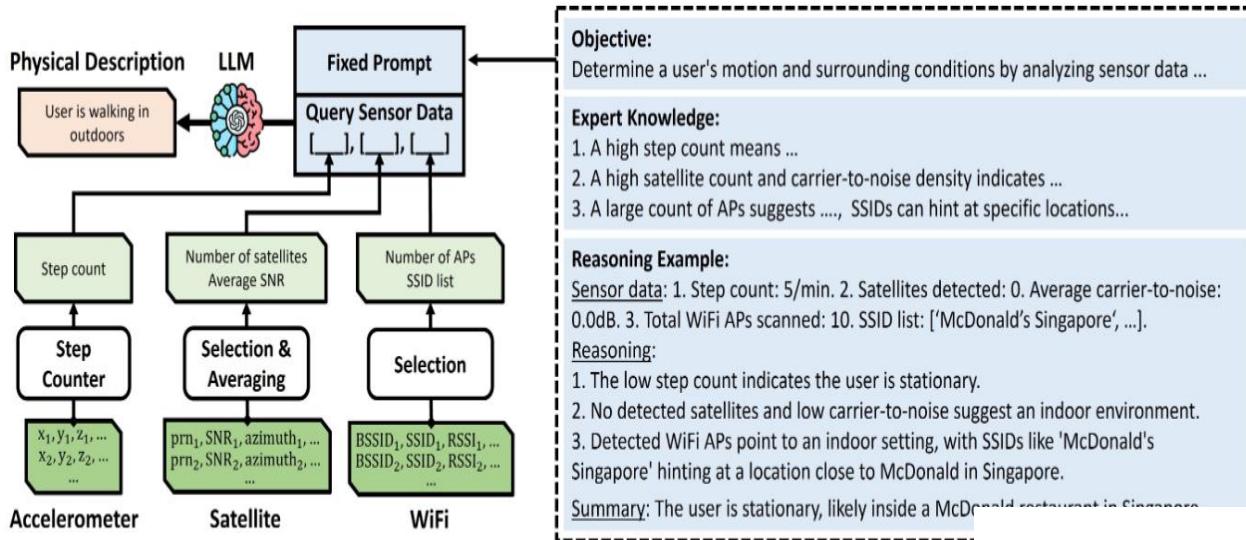


Figure 2: Overview of user activity sensing with LLMs.

2) detecção de batimentos cardíacos humanos onde dados digitalizados do eletrocardiograma (ECG) são utilizados para derivar a frequência cardíaca.

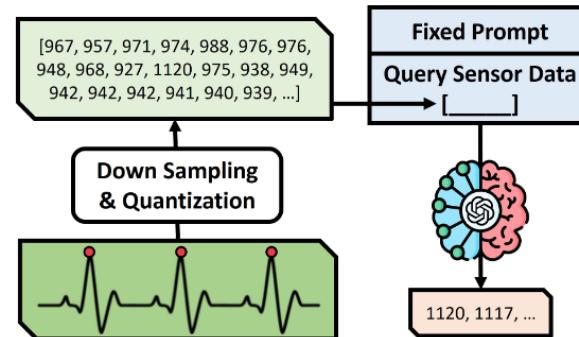


Figure 5: Overview of heartbeat detection with LLM.

Metodologia com duas aplicações ilustrativas em dois níveis:

1) **Detecção de atividade** do usuário onde sinais textualizados do acelerômetro do smartphone, dados de satélite e WiFi são analisados para discernir o movimento do usuário e as condições do ambiente.

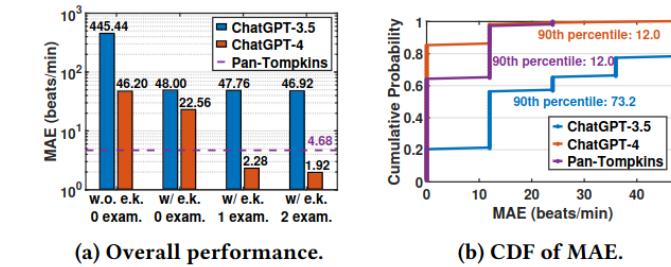


Figure 6: MAE of ChatGPT in heartbeat detection. (a) The 'e.k.' denotes expert knowledge, and 'exam.' refers to the reasoning examples provided.

LLM->Prompting

Penetrative AI: Making LLMs Comprehend the Physical World

"IA Penetrativa", um novo conceito relacionado à exploração do uso do conhecimento comum de modelos de linguagem grandes como base para realizar tarefas de percepção e intervenção no mundo real em sistemas ciber-físicos

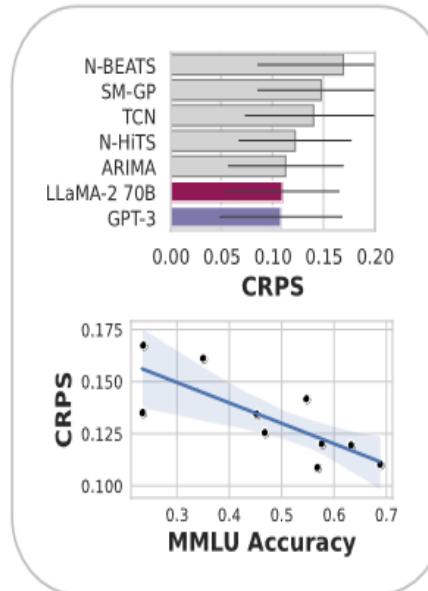
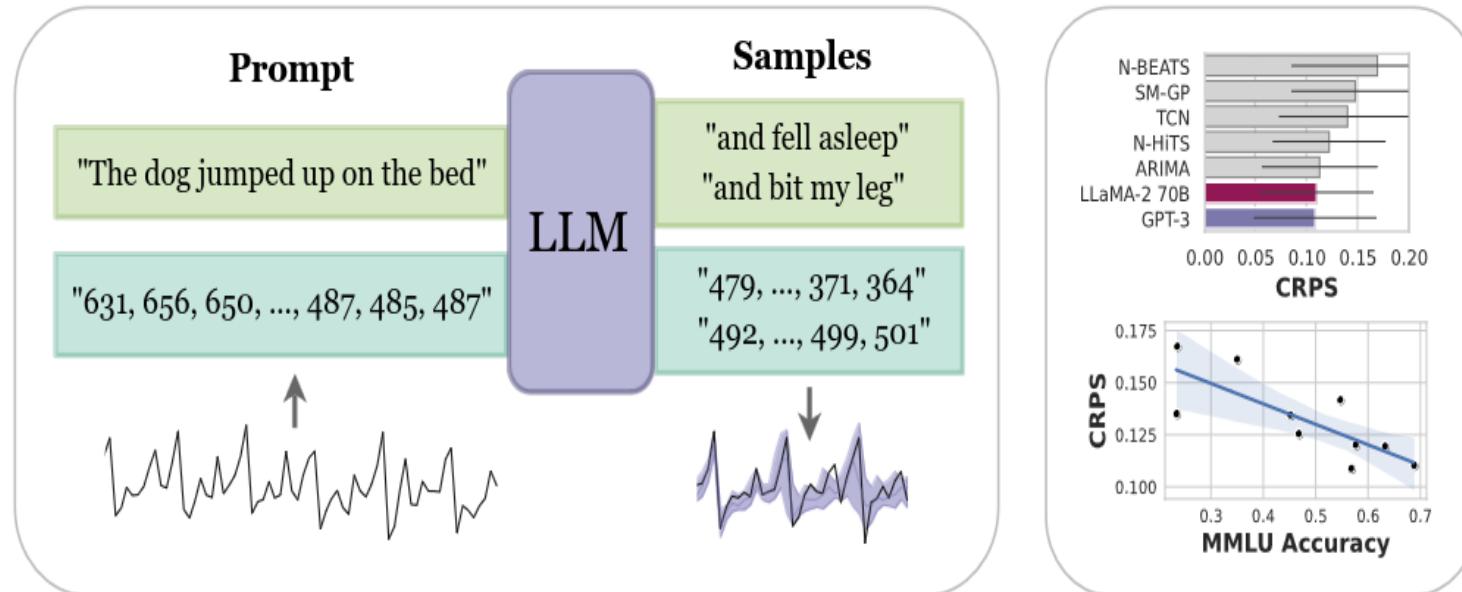
Penetrative AI: Making LLMs Comprehend the Physical World

<p>Input Sensor Data</p> <ol style="list-style-type: none">Step count: 105.0/min.Satellites detected: 0. Carrier-to-noise: 0.0dB.Total WiFi APs scanned: 31. SSID list: ['EastPointCity', 'WiFi.HK via EPCC', '3DG Jewellery', '3HK Pro3 5G', '5G私家寬頻', 'criticall01', 'CSL WiFi-Roam', 'H158-381_6c49', 'CENTALINE FREE WIFI', 'MiHome', 'SASA_LIVE', '_BWSHOP_', ...] <p>Reasoning:</p> <ol style="list-style-type: none">Step count: 105.0/min. The high step count indicates that the user is likely walking, suggesting movement and active state.Satellites detected: 0. Carrier-to-noise: 0dB. The absence of detected satellites and a carrier-to-noise ratio of 0 dB strongly suggest that the user is in an indoor environment, as satellite signals are usually weak or non-existent indoors.Total WiFi APs scanned: 31. The detection of 31 WiFi Access Points supports the inference of an indoor environment. Modern buildings usually have a high density of WiFi networks. Analyzing the SSID list, such as 'EastPointCity', 'WiFi.HK via EPCC', '3DG Jewellery', etc., it is reasonable to speculate that the user might be inside a shopping mall or a commercial area, possibly in Hong Kong, considering SSIDs like 'WiFi.HK via EPCC' and '3HK Pro3 5G'. 'EastPointCity' could be the name of a specific location or a mall. <p>Summary:</p> <p>The user is likely walking inside a commercial building or a shopping mall, possibly located in Hong Kong or a region where there is a presence of Hong Kong-based telecom and WiFi services. The active and diverse WiFi networks and the absence of satellite signals confirm the indoor setting, and the high step count confirms the walking motion.</p> <p>Response from ChatGPT-4</p> <p>Input Sensor Data</p>  <p>Shopping Mall, Hong Kong</p> <p>Lab, Los Angeles</p> <p>Airport, Singapore</p> <p>Input Sensor Data</p>  <p>Input Sensor Data</p>  <p>Input Sensor Data</p> <p>Input Sensor Data</p>  <p>Sports Field, Singapore</p> <p>Input Sensor Data</p> <p>Input Sensor Data</p> <p>Summary: The user is stationary and is likely inside a building on the UCLA campus or in its vicinity.</p> <p>Summary: The user is stationary, and they are likely inside or close to Changi Airport in Singapore.</p> <p>Summary: The user is likely walking in an outdoor environment.</p>

Figure 3: Response examples of ChatGPT-4 for activity sensingⁱⁱ.

LLM->Prompting

Large Language Models Are Zero Shot Time Series Forecasters [link](#)



Método: Os autores propõem técnicas para converter dados numéricos em tokens de texto e vice-versa para previsões de LLM.

[codigo github](#)

LLM->Prompting



Table 2: Papers in the Mobility, General, Table, Energy, and IoT Domains

Date	Paper	Domain	Task	LLM
11 Sep 2022	Leveraging Language Foundation Models for Human Mobility Forecasting	Mobility	Forecasting	BERT, BoBERTa, GPT-2, etc.
20 Sep 2022	PromptCast: A New Prompt-based Learning Paradigm for Time Series Forecasting	General	Forecasting	BART, BigBird, RoBERTa, etc.
19 Oct 2022	TabLLM: Few-shot Classification of Tabular Data with Large Language Models	Table	Classification	T0, GPT-3
10 Jul 2023	Large Language Models as General Pattern Machines	General	Forecasting, Translation, etc.	GPT-3, PaLM
29 Aug 2023	Where Would I Go Next? Large Language Models as Human Mobility Predictors	Mobility	Forecasting	GPT-3.5
7 Oct 2023	Large Language Models for Spatial Trajectory Patterns Mining	Mobility	Anomaly Detection	GPT-3.5, GPT-4, Claude-2
11 Oct 2023	Large Language Models Are Zero-Shot Time Series Forecasters	General	Forecasting	GPT-3, LLaMA-2
26 Oct 2023	Utilizing Language Models for Energy Load Forecasting	Energy	Forecasting	BART, Bigbird, Pegasus
16 Feb 2024	Time Series Forecasting with LLMs: Understanding and Enhancing Model Capabilities	General	Forecasting	GPT-3.5, GPT-4, LLaMA-2
25 Feb 2024	LSTPrompt: Large Language Models as Zero-Shot Time Series Forecasters by Long-Short-Term Prompting	General	Forecasting	GPT-3.5, GPT-4
2 Mar 2024	Evaluating Large Language Models as Virtual Annotators for Time-series Physical Sensing Data	IoT	Classification	GPT-4 (creates a table)

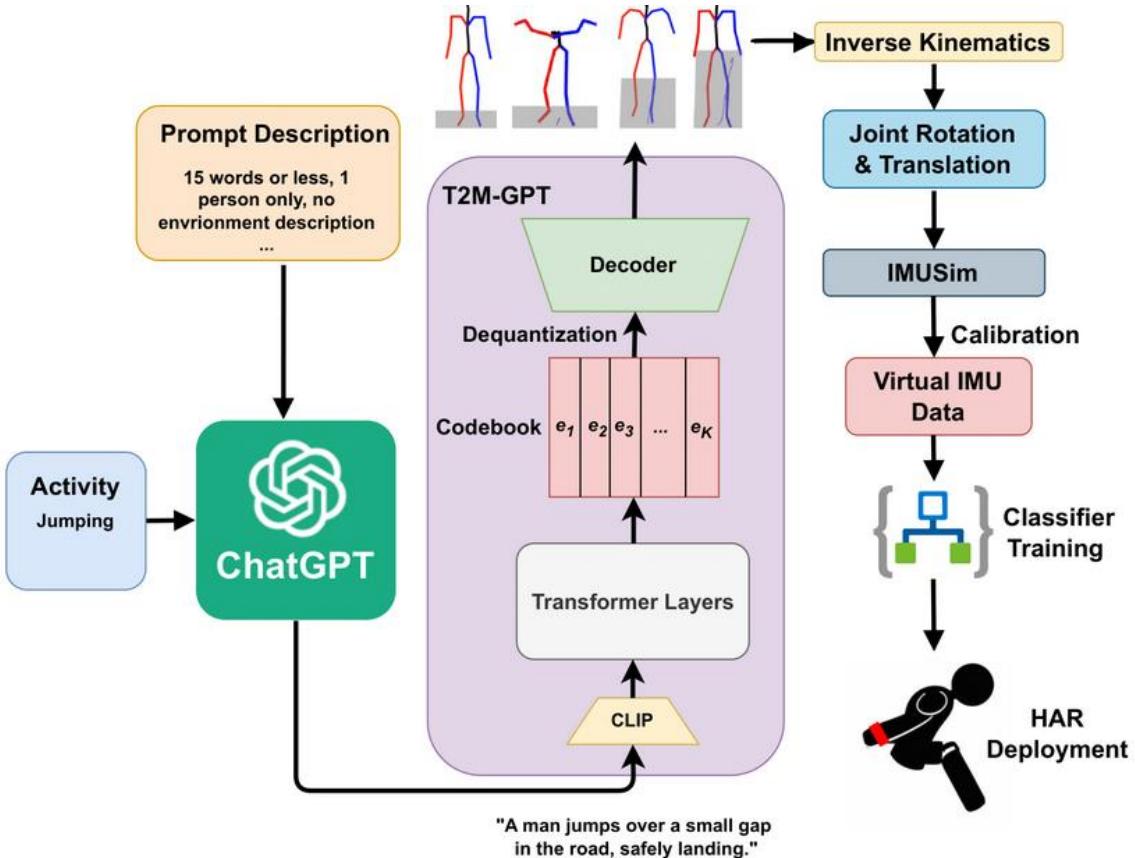
LLM->Prompting



Table 1: Papers in the Finance and Health Domains

Date	Paper	Domain	Task	LLM
30 Mar 2023	BloombergGPT: A Large Language Model for Finance	Finance	Sentiment Analysis, NER, QA, etc.	BloombergGPT
10 Apr 2023	The Wall Street Neophyte: A Zero-Shot Analysis of ChatGPT Over MultiModal Stock Movement Prediction Challenges	Finance	Forecasting	ChatGPT
24 May 2023	Large Language Models are Few-Shot Health Learners	Health	Classification, Regression	PaLM

Geração de dados
sintéticos



V1->Generating Virtual On-body Accelerometer Data from Virtual Textual Descriptions for Human Activity Recognition

[link](#)

IMUGPT 2.0: Language-Based Cross Modality Transfer for Sensor- Based Human Activity Recognition

[link](#)

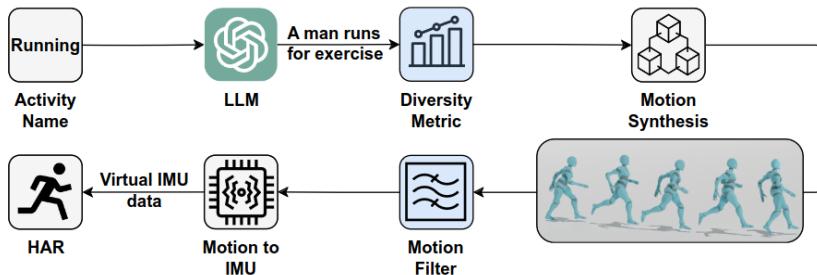


Fig. 1. Overview of the proposed language-based cross modality transfer system for sensor based human activity recognition. An LLM automatically generates textual descriptions of activities, which are then converted into motion sequences by a motion synthesis model. A novel motion filter then screens out incorrect sequences, retaining only relevant motion sequences for virtual IMU data extraction. A new diversity metric measures shifts in the distribution of generated textual descriptions and motion sequences, allowing for the definition of a stopping criterion that controls when data generation should be stopped for most effective and efficient processing and best downstream activity recognition performance.

IMUGPT 2.0: Language-Based Cross Modality Transfer for Sensor- Based Human Activity Recognition

[link](#)

codigo

	RealWorld	PAMAP2	USC-HAD	HAD-AW	MyoGym
Random Forest Classifier					
GPT-3.5	79.70 ± 0.38	69.20 ± 0.29	49.72 ± 0.67	48.98 ± 0.11	47.93 ± 0.58
GPT-4	79.41 ± 0.29	67.73 ± 0.70	49.07 ± 0.27	48.61 ± 0.20	47.64 ± 0.17
LLaMa 2	79.02 ± 1.01	67.82 ± 0.30	49.41 ± 0.28	48.40 ± 0.08	47.95 ± 0.25
Palm 2 (Bard)	78.79 ± 0.65	68.42 ± 0.43	49.06 ± 0.31	49.36 ± 0.05	47.91 ± 0.25
Gemini	78.93 ± 0.17	67.53 ± 0.10	49.65 ± 0.51	48.70 ± 0.06	47.98 ± 0.22
Real Data	71.53 ± 1.07	67.08 ± 0.53	47.23 ± 0.08	52.77 ± 0.07	46.61 ± 0.09
Deep ConvLSTM					
GPT-3.5	82.25 ± 0.32	75.16 ± 0.82	61.44 ± 0.49	51.98 ± 0.28	49.46 ± 0.88
GPT-4	80.20 ± 0.88	73.59 ± 0.75	61.43 ± 0.59	51.11 ± 0.28	46.85 ± 0.28
LLaMa 2	82.33 ± 0.34	74.12 ± 0.88	60.93 ± 0.39	50.94 ± 0.17	48.78 ± 0.45
Palm 2 (Bard)	81.74 ± 0.48	74.66 ± 0.96	61.17 ± 0.44	51.03 ± 0.11	49.00 ± 0.39
Gemini	80.86 ± 0.67	72.76 ± 0.37	61.02 ± 0.15	51.44 ± 0.49	48.82 ± 0.31
Real Data	77.79 ± 0.85	69.26 ± 1.07	63.35 ± 0.67	56.16 ± 0.38	50.69 ± 0.61
Deep ConvLSTM with self attention					
GPT-3.5	80.47 ± 0.49	73.45 ± 0.85	59.12 ± 0.62	47.67 ± 0.73	47.35 ± 0.41
GPT-4	80.31 ± 0.36	73.68 ± 1.26	57.97 ± 0.68	47.64 ± 0.50	46.03 ± 0.29
LLaMa 2	80.36 ± 0.67	73.77 ± 0.85	59.09 ± 0.89	48.75 ± 1.07	47.58 ± 0.47
Palm 2 (Bard)	80.77 ± 0.72	73.23 ± 0.92	58.66 ± 1.01	47.03 ± 0.90	47.21 ± 0.56
Gemini	80.82 ± 0.61	72.89 ± 0.34	58.84 ± 0.33	47.99 ± 0.06	47.10 ± 0.28
Real Data	77.50 ± 0.76	64.36 ± 0.52	61.82 ± 0.82	56.51 ± 0.19	50.63 ± 0.88

4.2.3 *Results.* Table 2 shows the downstream performance when different LLMs are used in IMUGPT for virtual IMU data generation. Overall, we find that the downstream performance is best when GPT-3.5 is used to generate



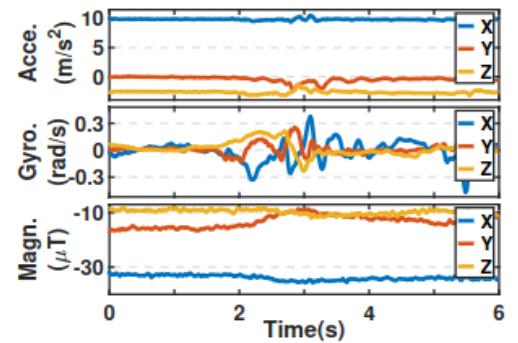
LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications



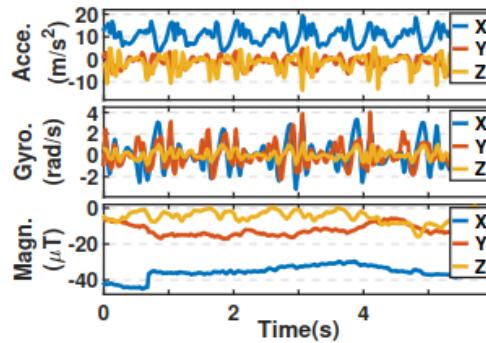
LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications

LIMU-BERT

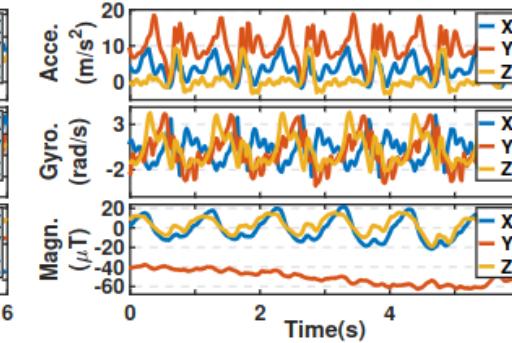
SenSys'21, November 15–17, 2021, Coimbra, Portugal



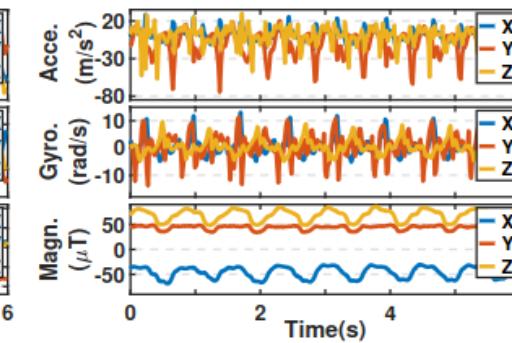
(a) Still, Bag



(b) Walk, Bag



(c) Walk, Pocket



(d) Run, Pocket

Figure 1: IMU measurements of daily activities with different device placements.

[Link artigo](#)

[Link código](#)

Resumo de LIMU-BERT

Problema: A maioria dos modelos de sensoriamento IMU exige muitos dados rotulados para serem treinados, o que é caro e trabalhoso.

Solução: LIMU-BERT, um modelo de aprendizado de representação que pode usar dados IMU não rotulados para extrair recursos generalizados.

Benefícios:

- Modelos mais leves e fáceis de implantar em dispositivos móveis.
- Melhor desempenho com menos dados rotulados.

Avaliação:

- LIMU-BERT supera as abordagens existentes em duas aplicações de sensoriamento IMU.

Detalhes:

- O LIMU-BERT adapta o modelo BERT para dados IMU usando técnicas especiais.
- O modelo foi treinado e avaliado em quatro conjuntos de dados abertos.
- O LIMU-BERT oferece um método eficiente para melhorar o desempenho de modelos de sensoriamento IMU com menos dados rotulados.

LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications

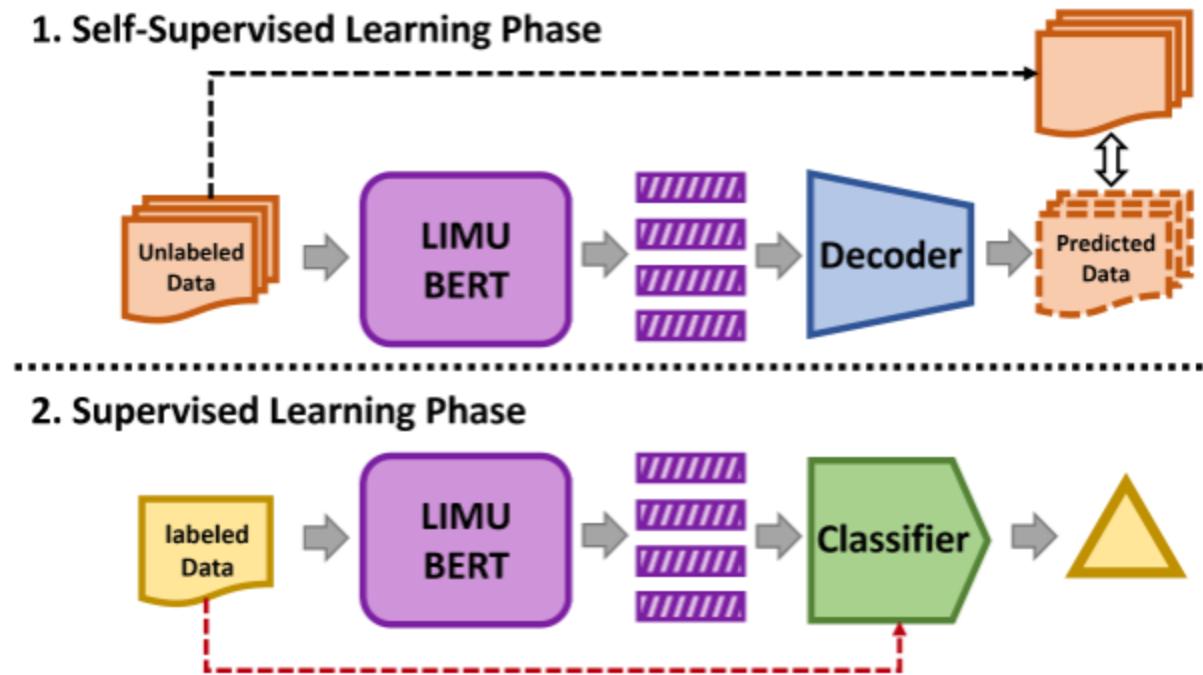


Figure 2: Framework overview.

LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications

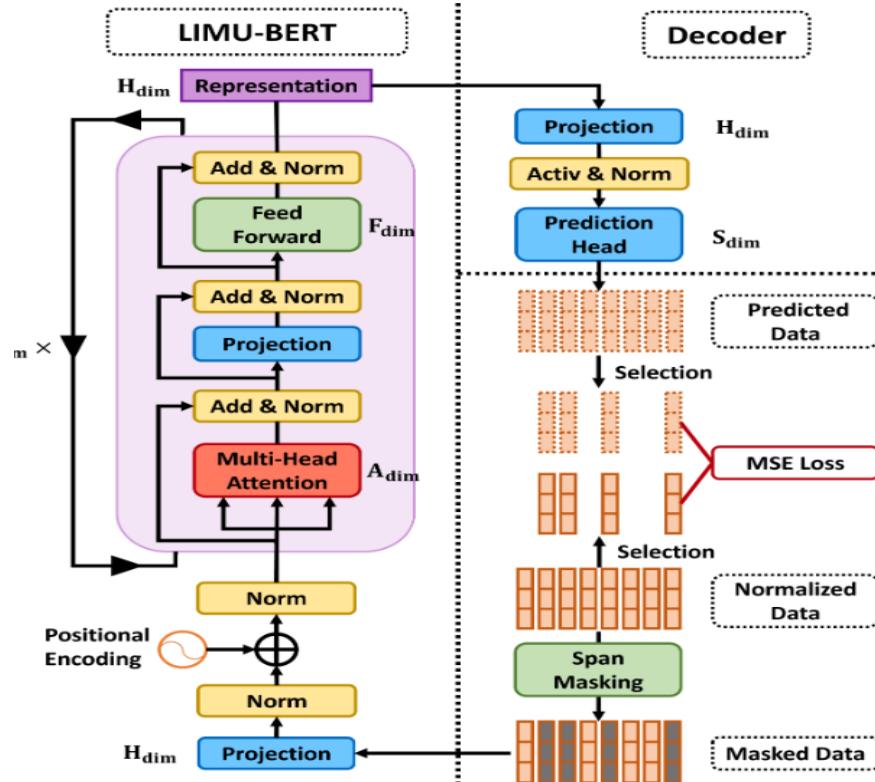


Figure 3: Self-supervised training workflow.

LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications

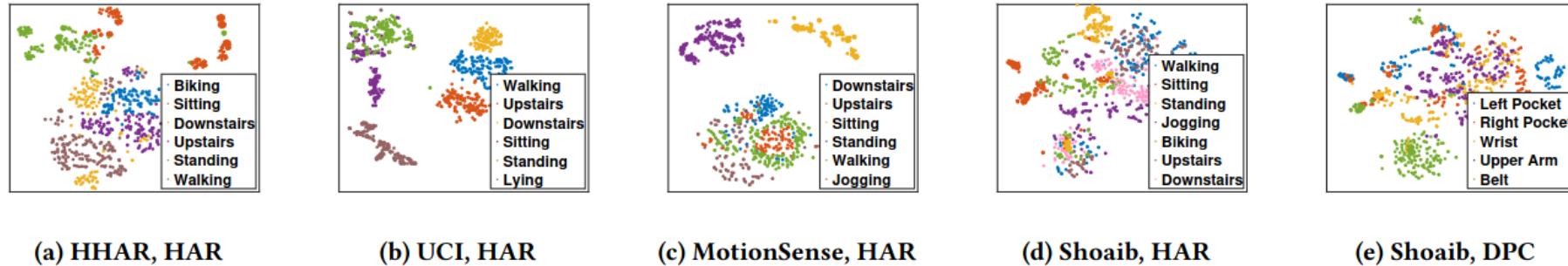


Figure 7: Representation visualization with t-SNE.

Table 4: Performance comparison of classifiers.

Application		HAR		HAR		HAR		HAR		DPC		Average	
Dataset	HAR	HHAR	UCI	MotionSense	Shoaib	Shoaib	-	Shoaib	Shoaib	-	-	-	-
Metric	Acc	F1											
LIMU-CNN	0.952	0.946	0.883	0.882	0.895	0.858	0.849	0.850	0.884	0.884	0.893	0.884	
LIMU-ATTN	0.928	0.923	0.915	0.913	0.909	0.874	0.809	0.810	0.812	0.811	0.875	0.866	
LIMU-LSTM	0.953	0.949	0.913	0.915	0.913	0.880	0.890	0.891	0.921	0.921	0.918	0.911	
LIMU-GRU	0.964	0.962	0.924	0.923	0.927	0.899	0.900	0.899	0.920	0.921	0.927	0.921	

Human Activity Recognition Transformer (HART) ❤

Human Activity Recognition Transformer (HART) ❤

- Arquitetura de **Transformer** leve, adaptada a sensores específica do domínio dos IMUs incorporados em dispositivos móveis. Nossos experimentos em tarefas de RAH com vários conjuntos de dados públicos disponíveis mostram que o HART usa menos Operações de Ponto Flutuante por Segundo (FLOPS) e parâmetros.
- apresentamos avaliações em várias arquiteturas em seu **desempenho em ambientes heterogêneos** e mostramos que nosso modelo pode generalizar melhor em diferentes dispositivos de detecção ou em **posições do corpo diferentes**.

[Artigo 1](#)

[Artigo 2](#)

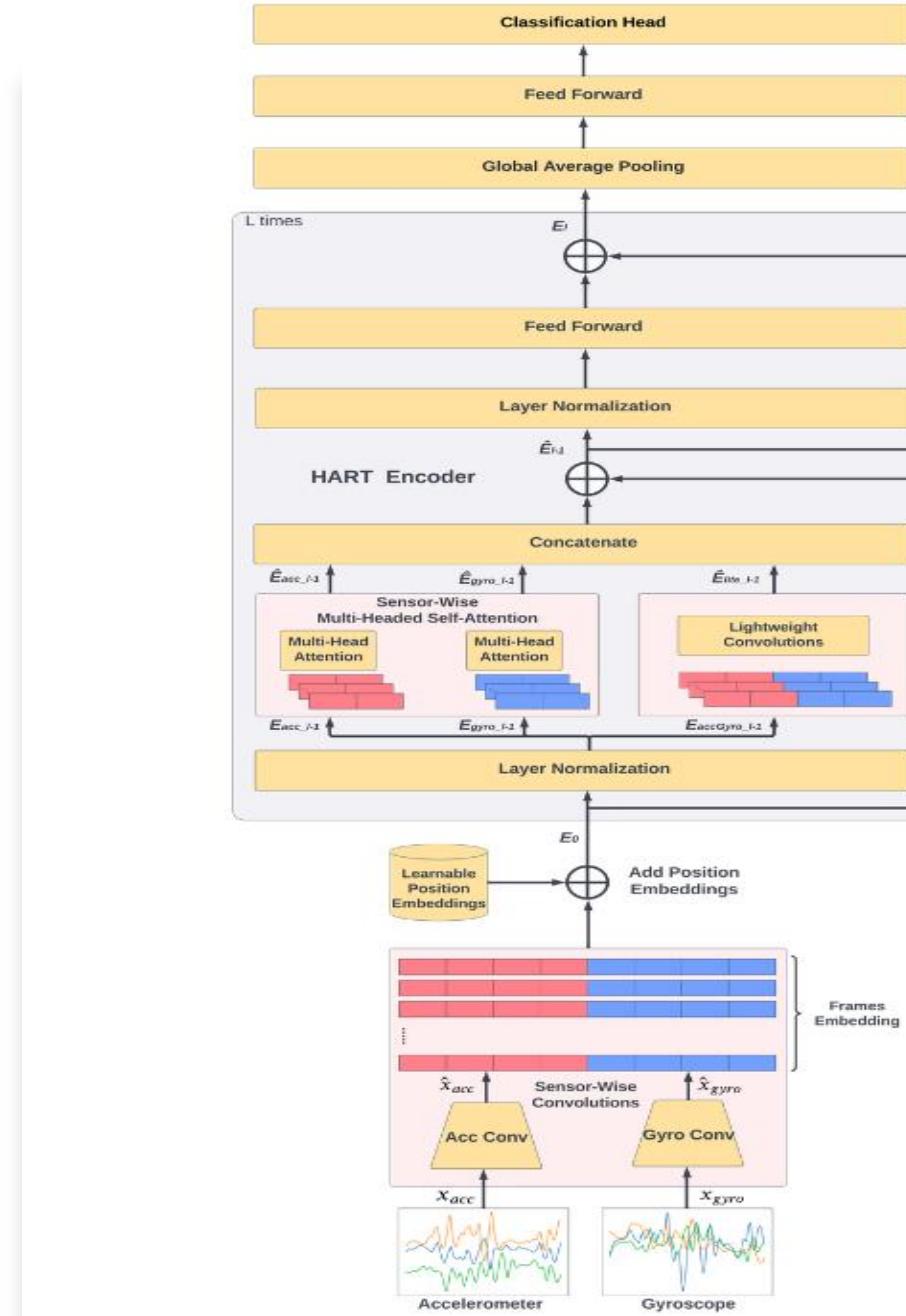
[Artigo 3](#)

Human Activity Recognition Transformer (HART)



[Codigo github](#)

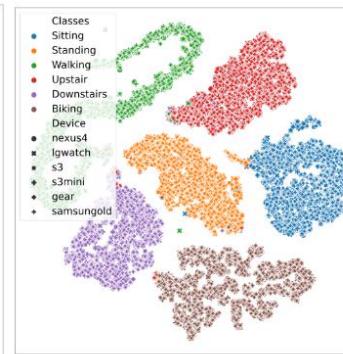
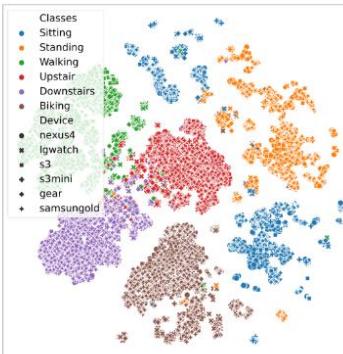
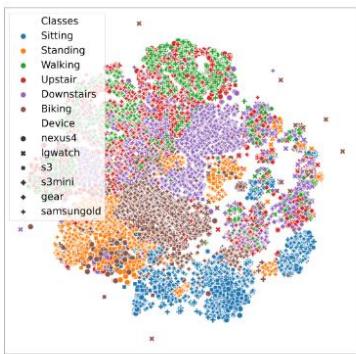
LIMU



Human Activity Recognition Transformer (HART) ❤️

[Codigo github](#)

Lightweight Transformers for Human Activity Recognition on Mobile Devices - Preprint Version

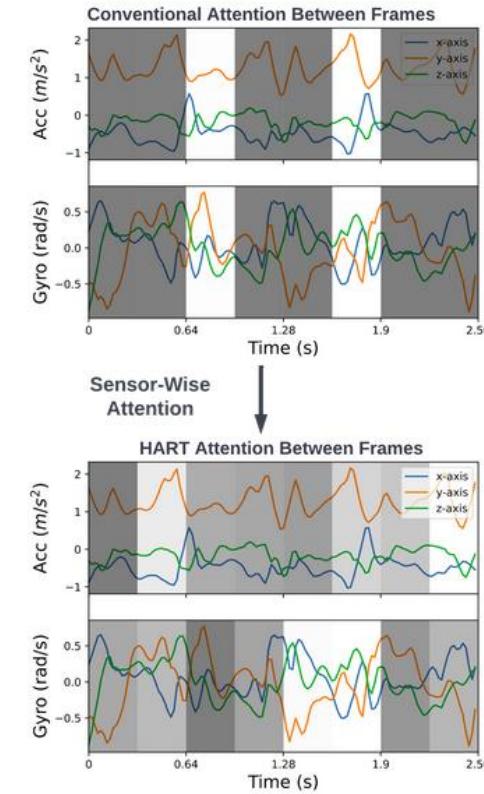


5. Results

The table below shows the results obtained with HART and MobileHART on our training and evaluation pipeline with the 5 and combined datasets.

Architecture	UCI	MotionSense	HHAR	RealWorld	SHL	Combined
HART	94.49	98.20	97.36	94.88	79.49	85.61
MobileHART	97.20	98.45	98.19	95.22	81.36	86.74

The image below shows the result of our sensor-wise attention compared against conventional attention:



InfoGCN: Representation Learning for Human Skeleton- based Action Recognition⁺。

+ .
o

InfoGCN: Representation Learning for Human Skeleton-based Action Recognition

[Link1](#)

[Link2](#)

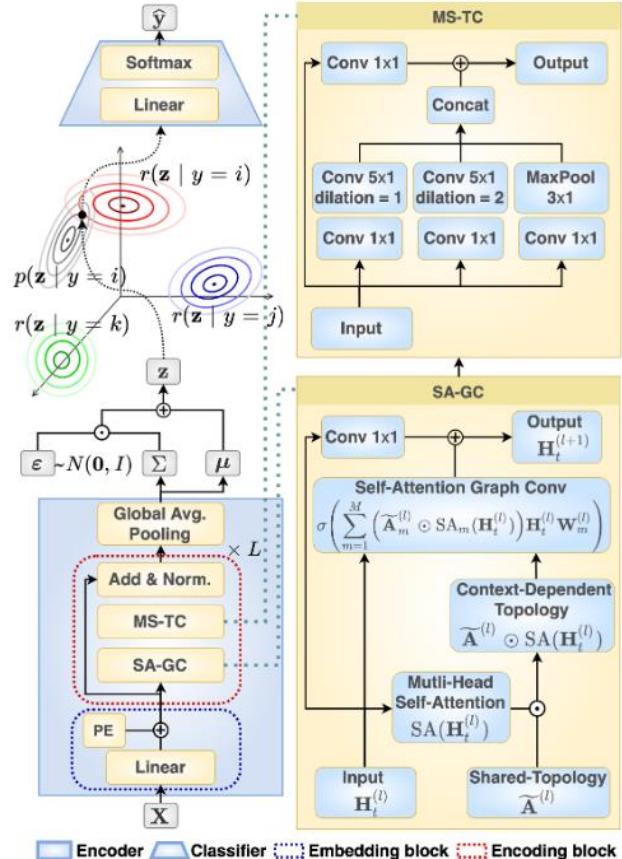


Figure 2. InfoGCN Architecture. We guide our neural architecture to learn the class conditional representation of skeleton-based action with the information bottleneck objective. The model is composed of an encoder and a classifier. The encoder with the SA-GC module captures context-dependent intrinsic joint topology to better represent action.

InfoGCN: Representation Learning for Human Skeleton-based Action Recognition

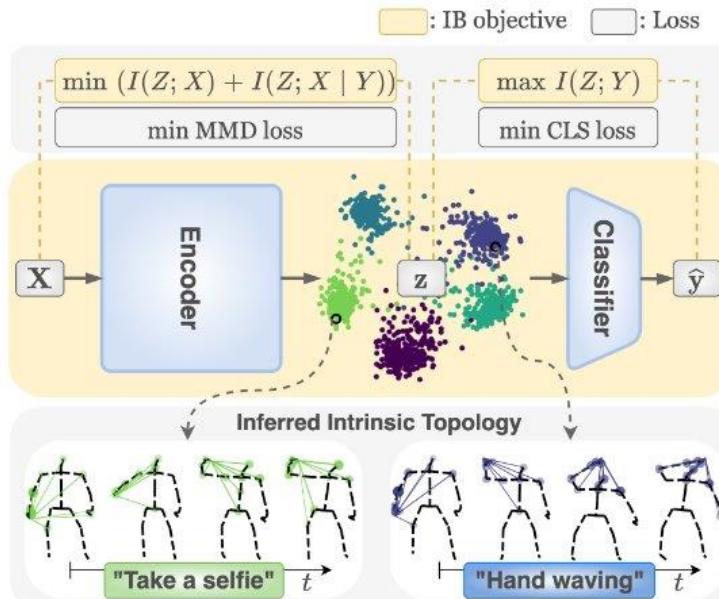


Figure 1. Conceptual diagram of InfoGCN. We propose an IB objective and a corresponding loss to guide our model to learn maximally informative representations for skeleton-based action recognition. The encoder infers an intrinsic topology of joints, which provides contextual information beyond physical connectivity. The colored lines on the bottom indicate inferred intrinsic topology, and the thickness represents the strength of the relation.

InfoGCN: Representation Learning for Human Skeleton-based Action Recognition

Methods	Acc (%)	
	X-Sub	X-View
ST-GCN [56]	81.5	88.0
AS-GCN [27]	86.8	94.2
2s-AGCN [44]	88.5	95.1
SGN [60]	89.0	94.5
DGNN [43]	89.9	96.1
ST-TR-agcn [39]	90.3	96.3
Shift-GCN [9]	90.7	96.5
DC-GCN+ADG [8]	90.8	96.6
PA-ResGCN-B19 [47]	90.9	96.0
DDGCN [22]	91.1	97.1
Dynamic GCN [58]	91.5	96.0
MS-G3D [33]	91.5	96.2
MST-GCN [7]	91.5	96.6
CTR-GCN [6]	92.4	96.8
Ours	93.0	97.1

Methods	Acc (%)	
	X-Sub	X-Set
SGN [60]	79.2	81.5
2S-AGCN [44]	82.9	84.9
ST-TR-agcn [39]	85.1	87.1
Shift-GCN [9]	85.9	87.6
DC-GCN+ADG [8]	86.5	88.1
MS-G3D [33]	86.9	88.4
PA-ResGCN-B19 [47]	87.3	88.3
Dynamic GCN [58]	87.3	88.6
MST-GCN [7]	87.5	88.8
CTR-GCN [6]	88.9	90.6
Ours (Joint)	85.1	86.3
Ours (Bone)	87.3	88.5
Ours (Joint + Bone)	88.5	89.7
Ours (4 ensemble)	89.4	90.7
Ours (6 ensemble)	89.8	91.2

Methods	Acc (%)
Lie Group [52]	74.2
Actionlet ensemble [54]	76.0
HBRNN-L [11]	78.5
Ensemble TS-LSTM [24]	89.2
AGC-LSTM [45]	93.3
Shift-GCN [9]	94.6
DC-GCN+ADG [8]	95.3
CTR-GCN [6]	96.5
Ours	97.0

Table 1. Comparative results on NTU RGB+D 60 [42] (*left*), NTU RGB+D 120 [30] (*middle*), and NW-UCLA [55] (*right*). We evaluate our model in terms of classification accuracy (%). The performance of baseline methods is based on their papers. Bold figures indicate the best value for each dataset. X-Sub, X-view, and X-Set represent cross-subject, cross-view, and cross-setup splits, respectively.

InfoGCN: Representation Learning for Human Skeleton-based Action Recognition

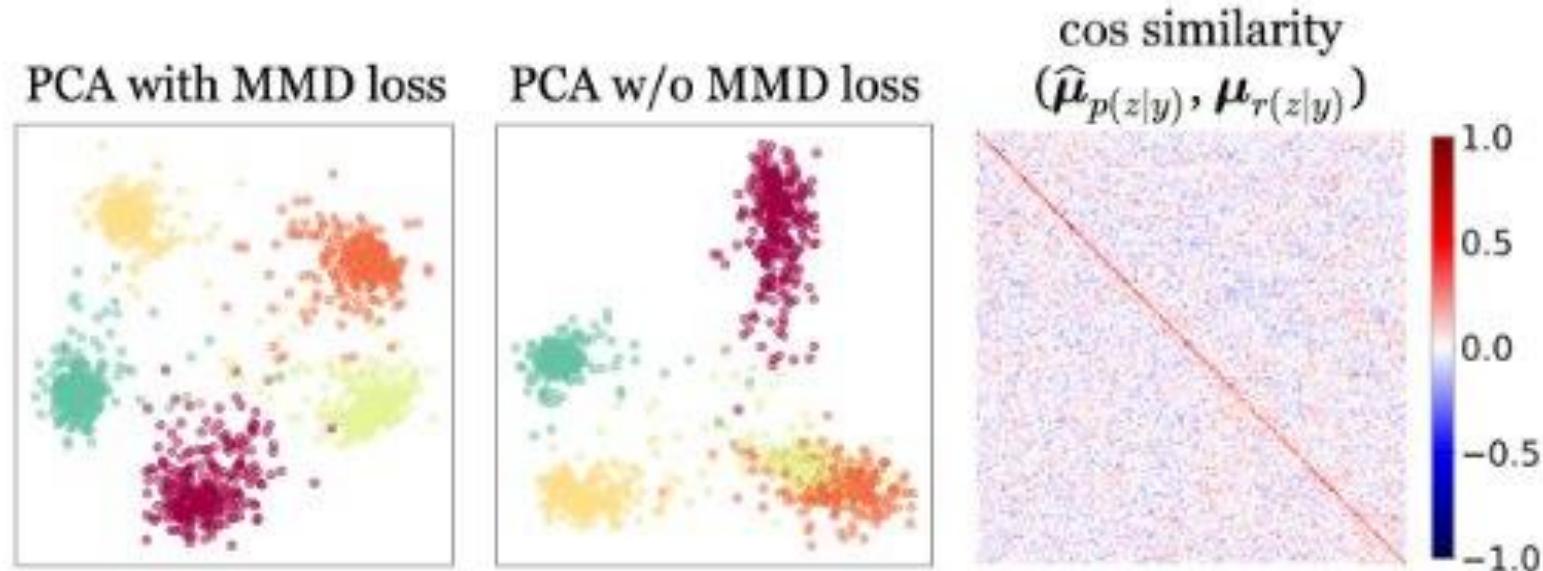


Figure 4. (*Left-middle*) PCA projection of latent representation to 2D when trained with or without MMD loss. We randomly select five action classes for visualization from NTU RGB+D 120 dataset. Different colors indicate different classes. (*Right*) cosine similarity between $\hat{\mu}_{p(z|y)}$ and $\mu_{r(z|y)}$. Each row and column indicates different classes.

+

•

◦

Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor

Data⁺◦

Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data

- Link

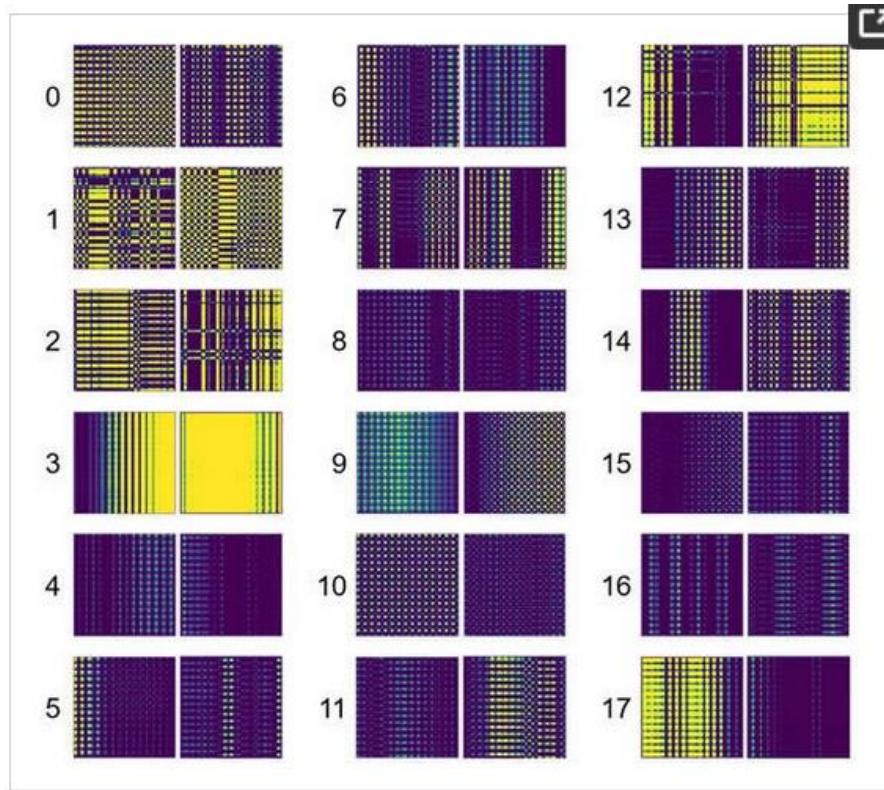


Figure 4. SBIDP example of 18 activities in the WISDM dataset.

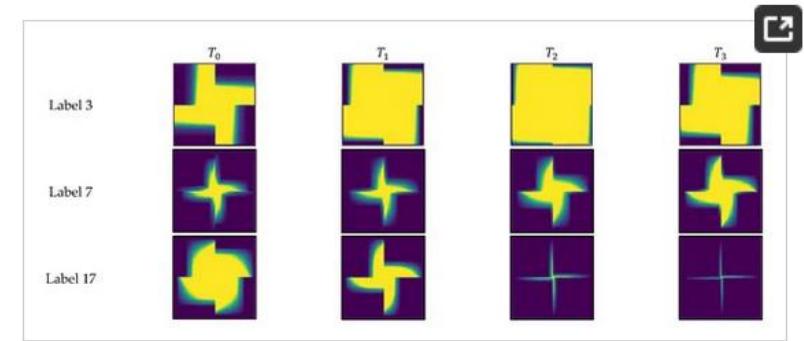
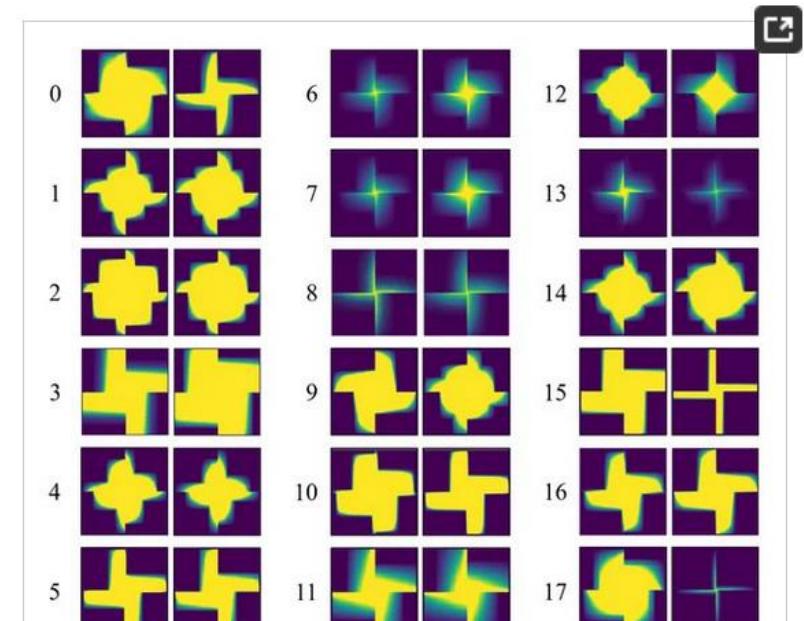


Figure 8. $BIDP_{E2}$ examples of 2nd enhanced $BIDP_{E1}$.



Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data

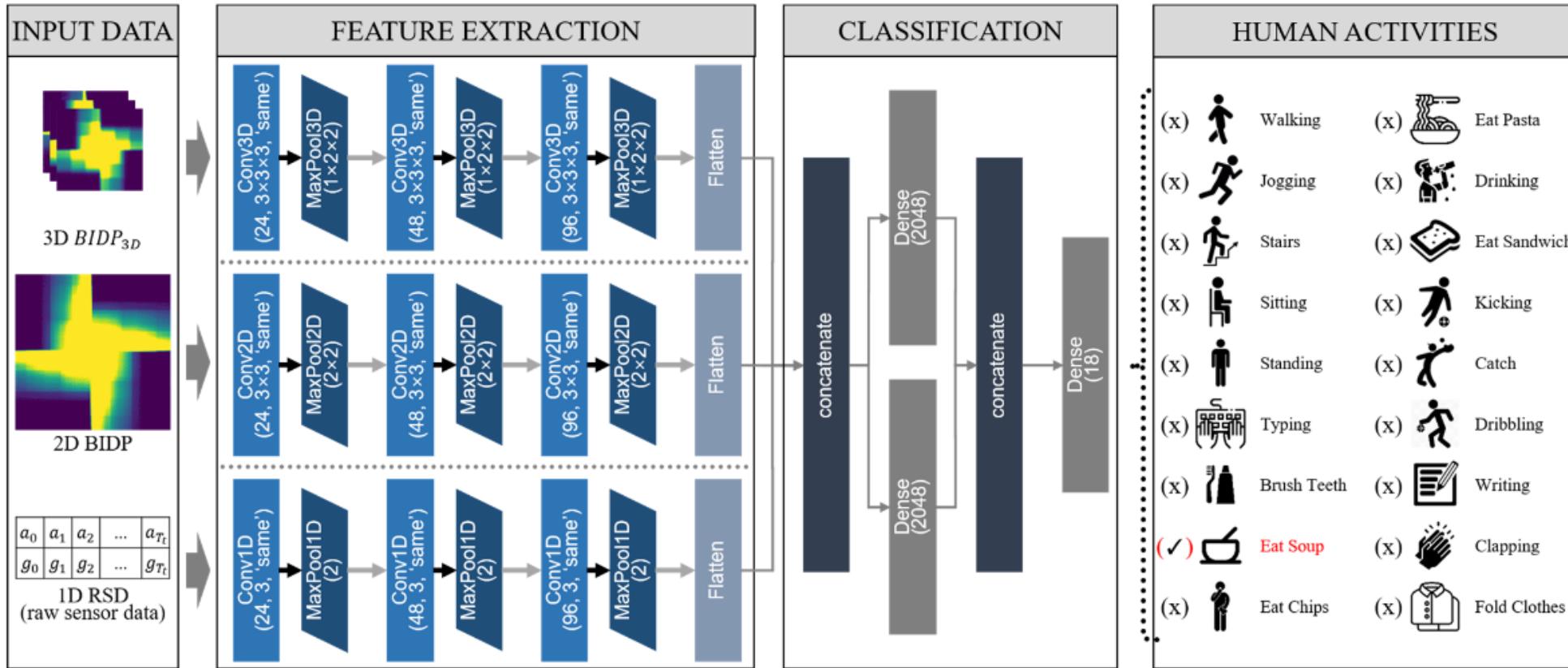


Figure 12. A Proposed multi-dimensional convolutional neural network.

Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data

Table 4. Evaluation of the proposed model compared with models based on RNN.

Ref.	Model	F1 Score (%)	Accuracy (%)
[31]	Tri-PSRNN	96.62	94.76
[31]	PSDRNN	94.01	93.06
[32]	LSTM-CNN	-	95.85
[33]	LSTM-RNN	95.40	96.40
[34]	Single-input CNN-GRU model A	92.42	92.03
[34]	Single-input CNN-GRU model B	94.50	94.71
[34]	Single-input CNN-GRU model C	92.55	92.37
[34]	Multi-input CNN-LSTM	95.55	95.45
[34]	Multi-input CNN-GRU	97.22	97.21
[35]	CNN-GRU-LSTM	98.52	98.51
-	Proposed model	98.00	98.15

+ · Single Input Single
◦ Head CNN-GRU-
LSTM Architecture
for
Recognition of
Human Activities⁺ .

[link](#)

Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities

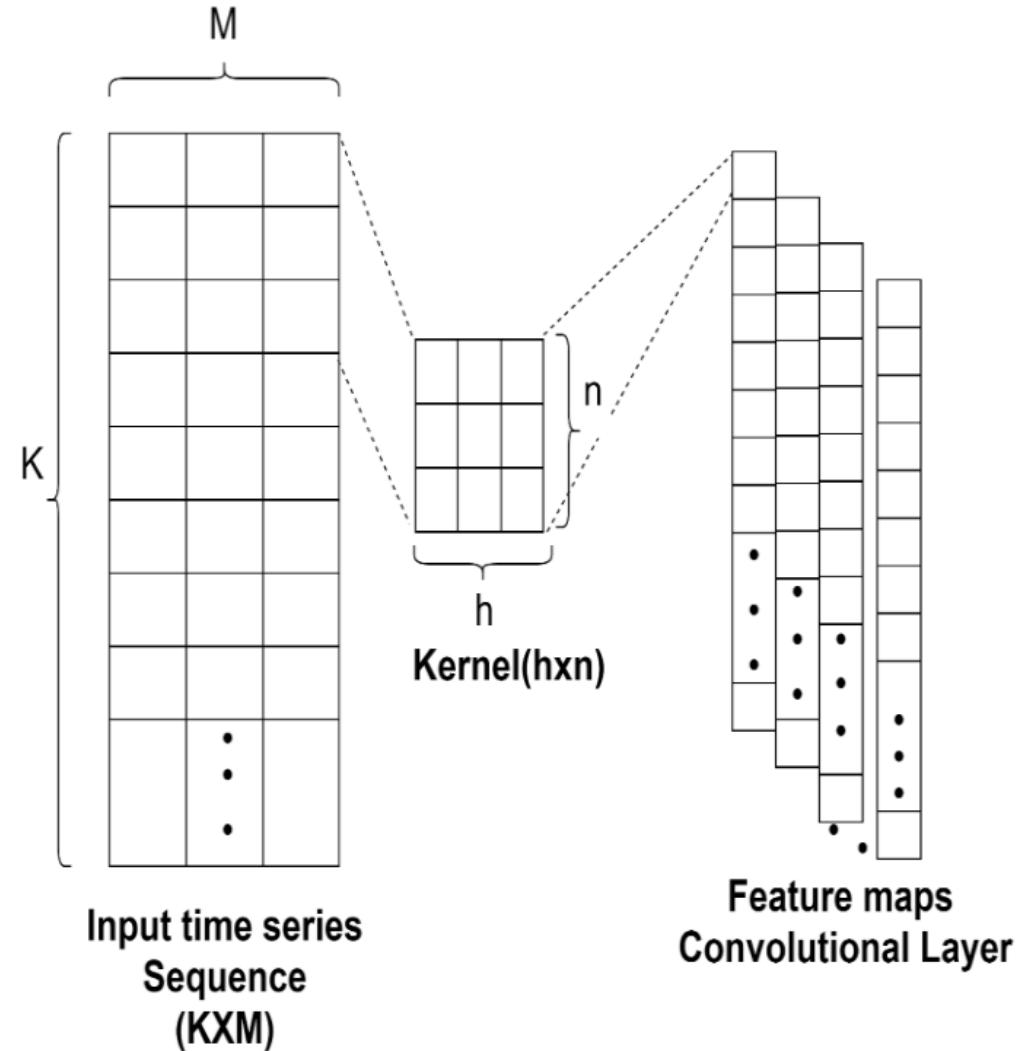


Figure 1. Operation of Convolutional layer on input data

Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities

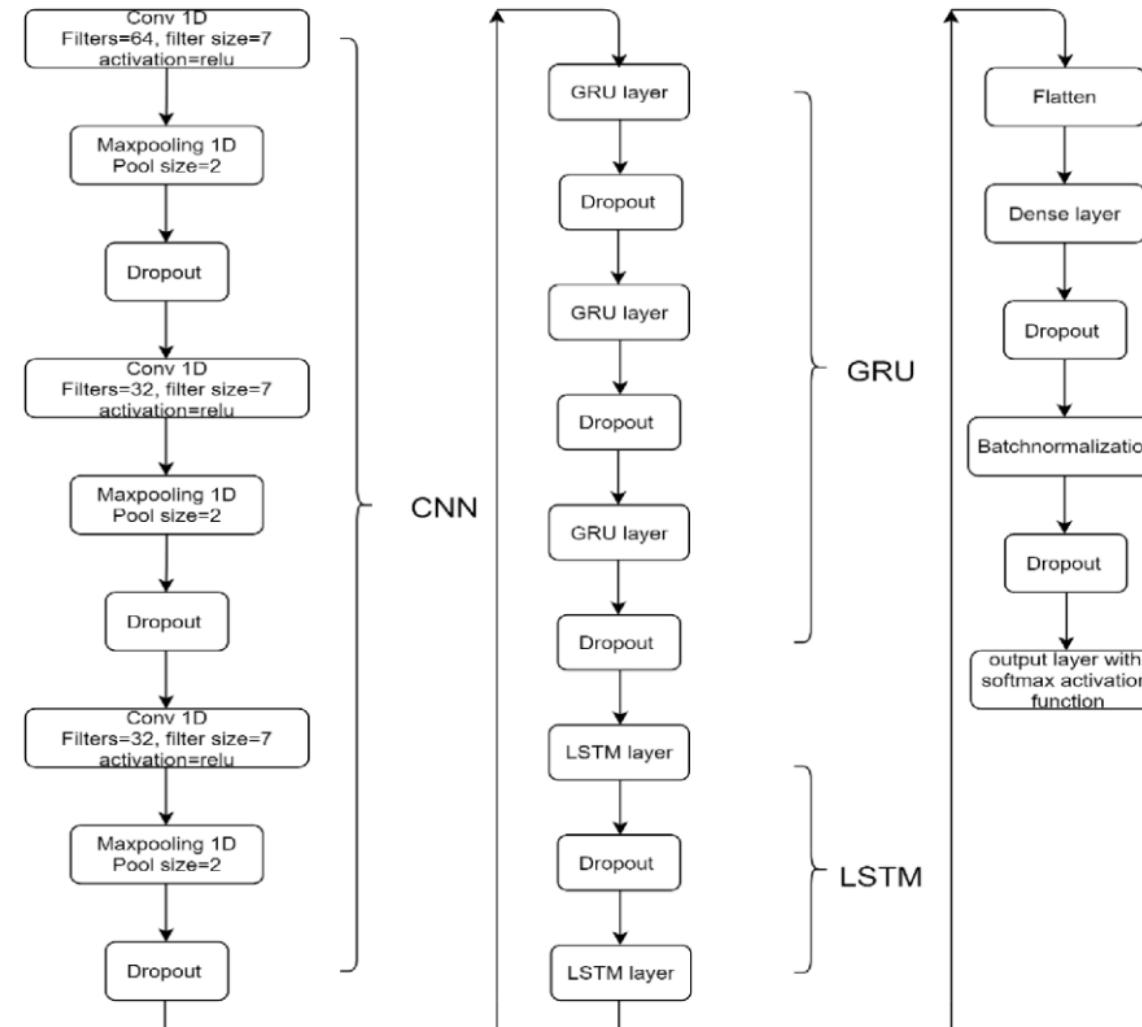


Figure 3 proposed Architecture CNN-GRU-LSTM for human activity recognition

Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities

- a. Este artigo propõe uma **arquitetura híbrida** de deep neural networks ao combinar os pontos fortes da CNN, GRU e LSTM para reconhecer atividades humanas tomando os dados brutos do sensor vestível com passos de pré-processamento insignificantes e sem um processo de extração de recursos feito à mão.
- b. Características locais são extraídas pela camada **CNN e GRU; LSTM** mantém as dependências temporais de longo prazo dessas características mapeadas, de modo que o modelo possa reconhecer dados diversos.
- c. O modelo foi testado em dois conjuntos de dados públicos disponíveis **UCI-HAR, WISDM e obteve precisão de 93,48% e 98,51%**, respectivamente.

DATASET:WIDSM, UCI

Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities

Table 2. Comparative performance of proposed model on UCI-HAR dataset

Frameworks	Accuracy (%)	F-1 score (%)
Res-LSTM [38]	91.6	91.5
CNN-LSTM [39]	92.13	-
Stacked LSTM [40]	93.13	-
CNN [41]	92.71	92.93
Single input CNN-GRU model A[42]	93.03	93.01
Single input CNN-GRU model B[42]	92.43	92.42
Proposed	93.48	93.5

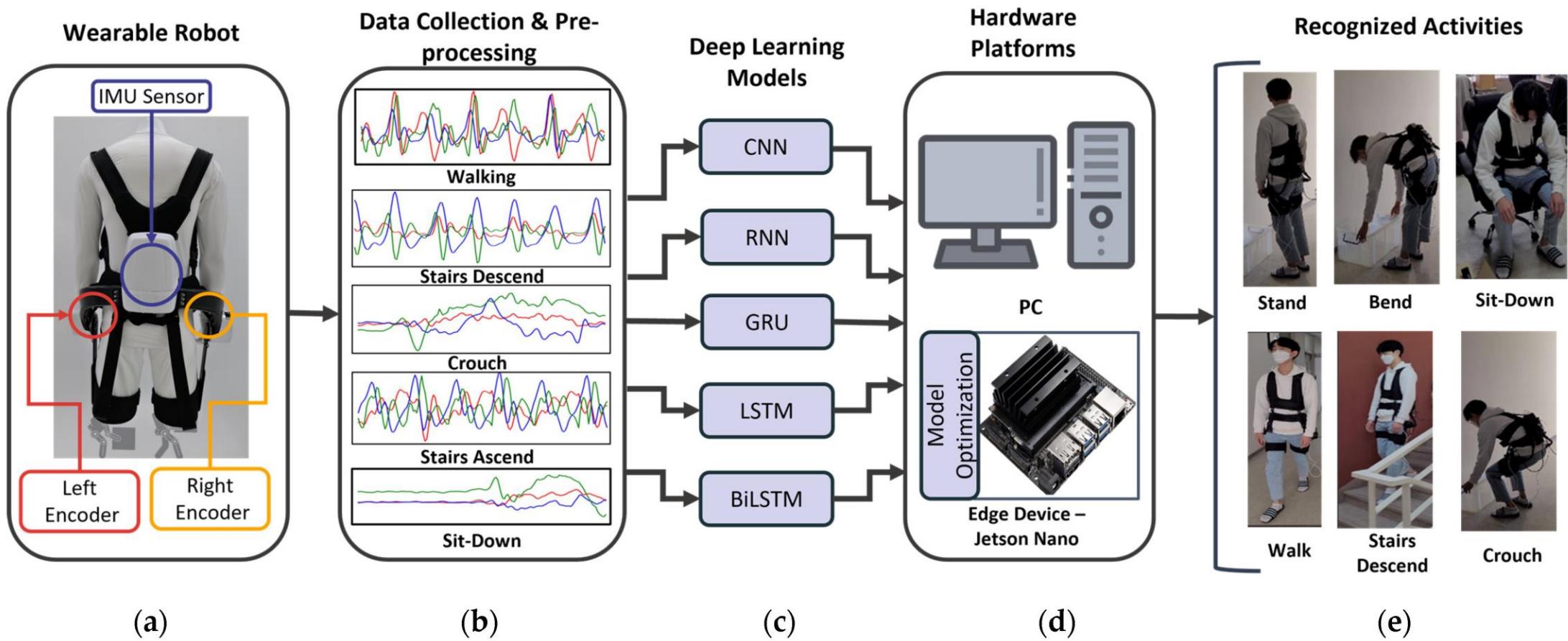
Table 3. Comparative performance of frameworks on WISDM dataset

Frameworks	Accuracy (%)	F-1 score
Statistical features and reweighted genetic algorithms [43]	94.02	-
CNN [44]	93.32	-
LSTM-CNN [45]	95.85	-
U-Net [46]	96.4	96.5
Single input CNN-GRU model A [42]	92.03	92.42
Single input CNN-GRU model B [42]	94.71	94.50
Single input CNN-GRU model C [42]	92.37	92.55
Multi input CNN-LSTM [42]	95.54	95.55
Multi input CNN-GRU [42]	97.21	97.22
Proposed	98.51	98.52

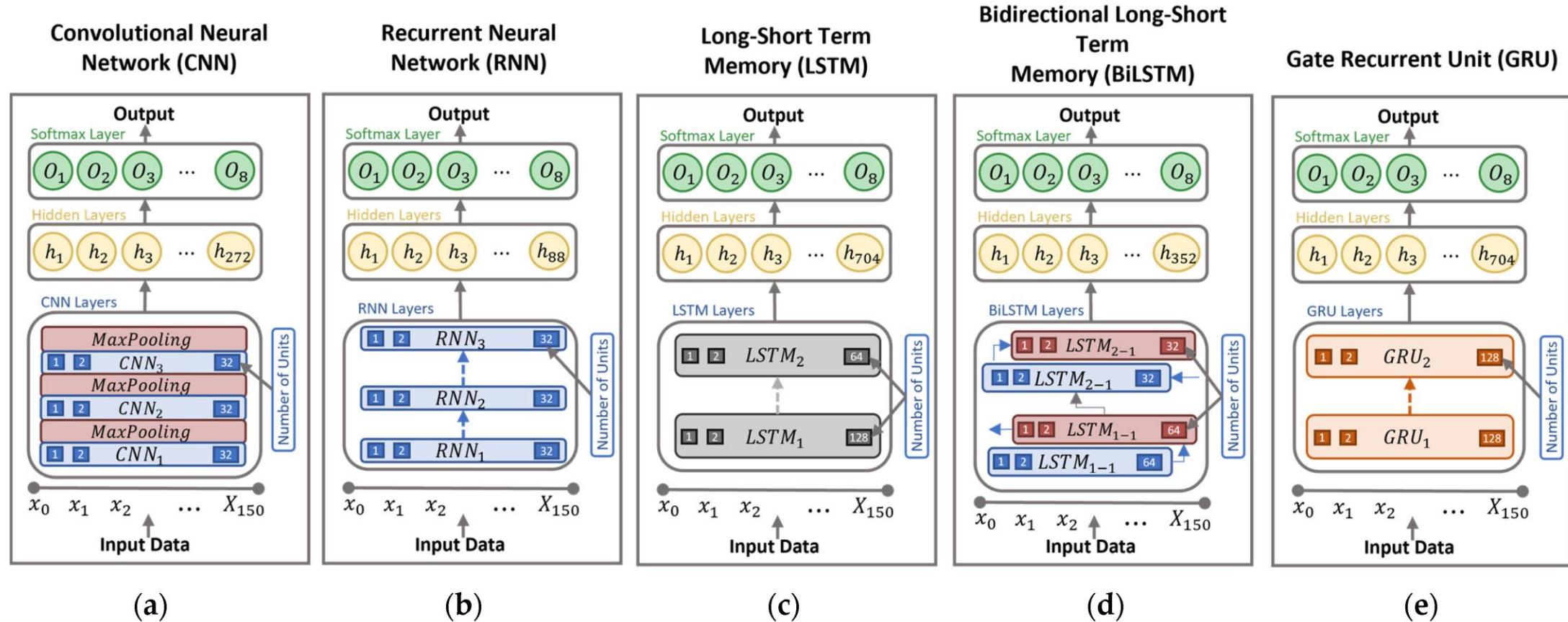
n-Activity-Recognition-using-CNN-and-LSTM-RNN

Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks

Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks



Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks



Real-Time Human Activity Recognition with IMU and Encoder Sensors in Wearable Exoskeleton Robot via Deep Learning Networks

[link](#)

testes realizados em PC e dispositivo de borda, bem como os testes em tempo real online. Foram comparados e validados a precisão usando o conjunto de dados do epoch e calculou-se o tempo de inferência para determinar os melhores modelos para o dispositivo de borda. Em seguida, três modelos selecionados foram incorporados no dispositivo Jetson Nano, validando o desempenho destes modelos por meio de testes em tempo real.

Text2Image: A Novel Signal- Encoding Technique for CNN-Based Human Activity Recognition

+

o

.

Iss2Image: A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition

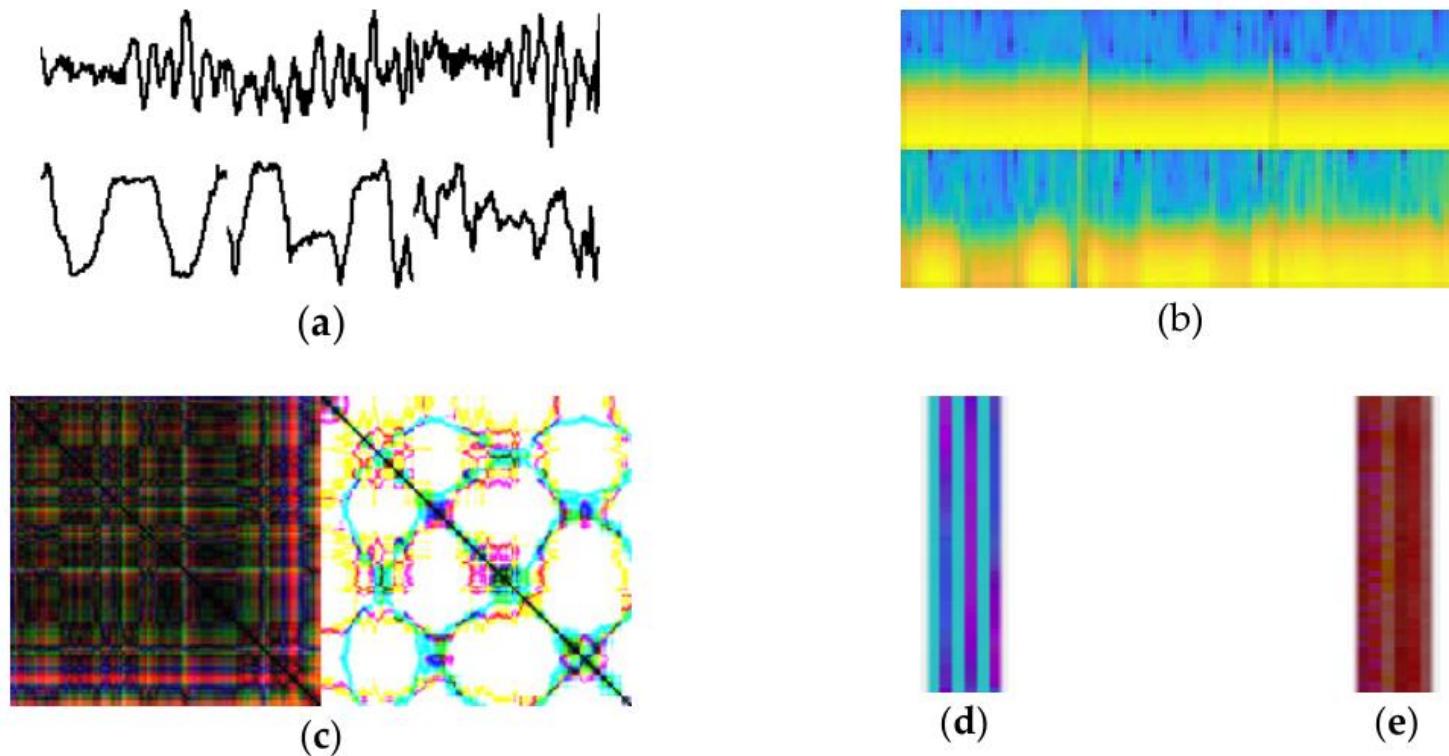


Figure 5. An illustration of activity images generated by (a) raw signal plot, (b) spectrogram, (c) recurrence plot, (d) multichannel, and (e) Iss2Image.

Iss2Image: A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition

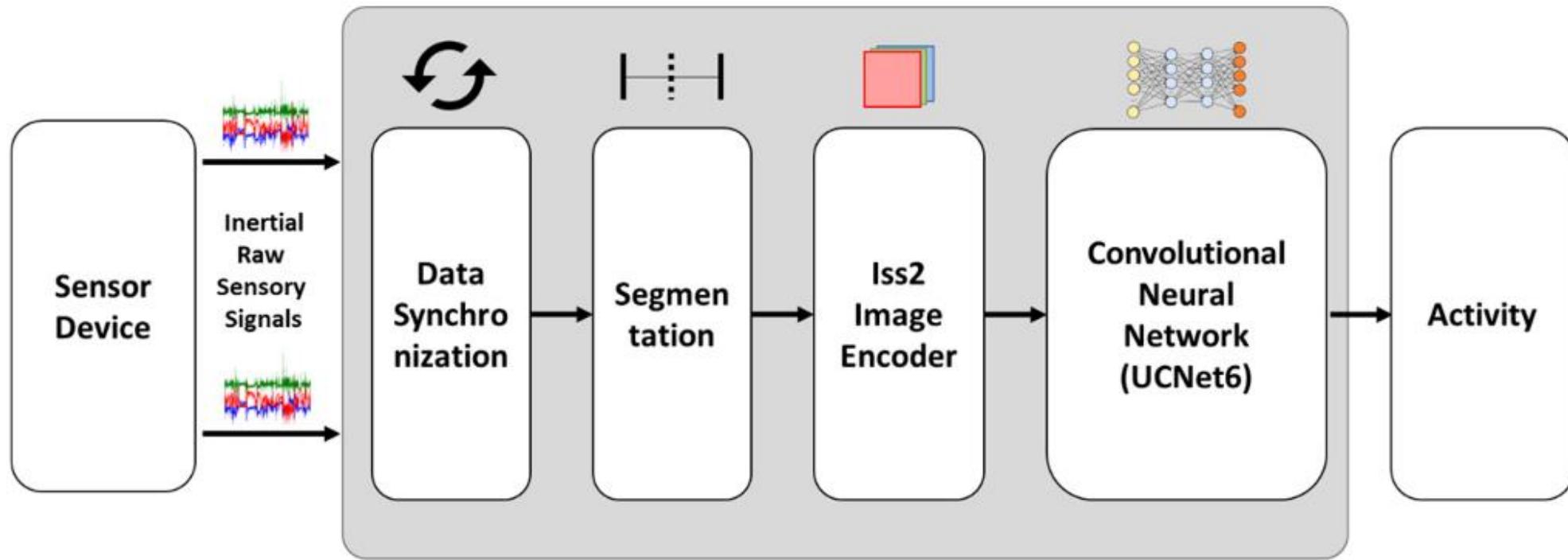


Figure 1. The workflow of our proposed method for human activity recognition; the main contributions are the data encoder and CNN model.

Iss2Image: A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition

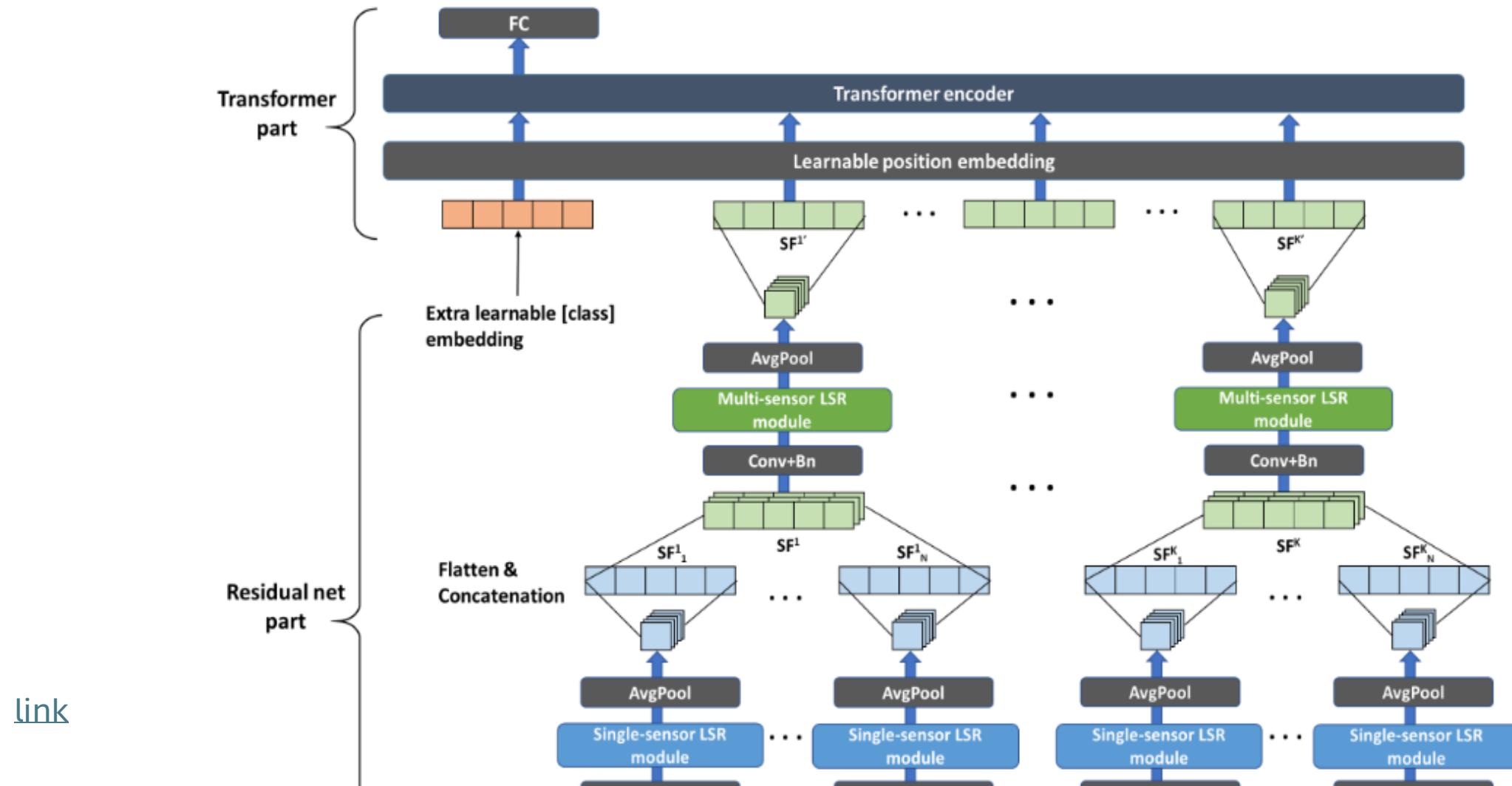
Table 3. Average recognition accuracy of our proposed method on the different datasets.

Dataset	Accuracy (%)
MobiAct	100.00
DaLiAc	98.90
UCI-HAR	97.11
UC-HAR	98.16

⁺UMSNet: An Universal Multi-sensor Network for Human Activity Recognition



UMSNet: An Universal Multi-sensor Network for Human Activity Recognition



UMSNet: An Universal Multi-sensor Network for Human Activity Recognition

Method comparison	$K = 6$		$K = 12$		$K = 24$	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
GRU[5]	0.878	0.876	0.9337	0.933	0.9187	0.9203
LSTM[16]	0.8455	0.8417	0.9145	0.9142	0.9346	0.9341
ResNet[27]	0.9707	0.9706	0.9701	0.9701	0.9714	0.9713
Efficientnet[31]	0.9783	0.9782	0.9714	0.9713	0.9736	0.9735
Regnet[27]	0.9785	0.9785	0.9813	0.9814	0.9836	0.9836
UMSNet-A(Ours)	0.9797	0.9798	0.983	0.9829	0.9852	0.9851
UMSNet-B(Ours)	0.9802	0.9802	0.9842	0.9842	0.9857	0.9856
UMSNet-C(Ours)	0.9804	0.9804	0.9844	0.9844	0.9872	0.9872

Table 4. Performance of different methods on MHEALTH dataset. K is the slice number of a time series data.

[link](#)



WIDEWALLPAPERS.NET

CONTRASTIVE LEARNING

Virtual Fusion with Contrastive Learning for Single Sensor-based Activity Recognition



Virtual Fusion with Contrastive Learning for Single Sensor-based Activity Recognition

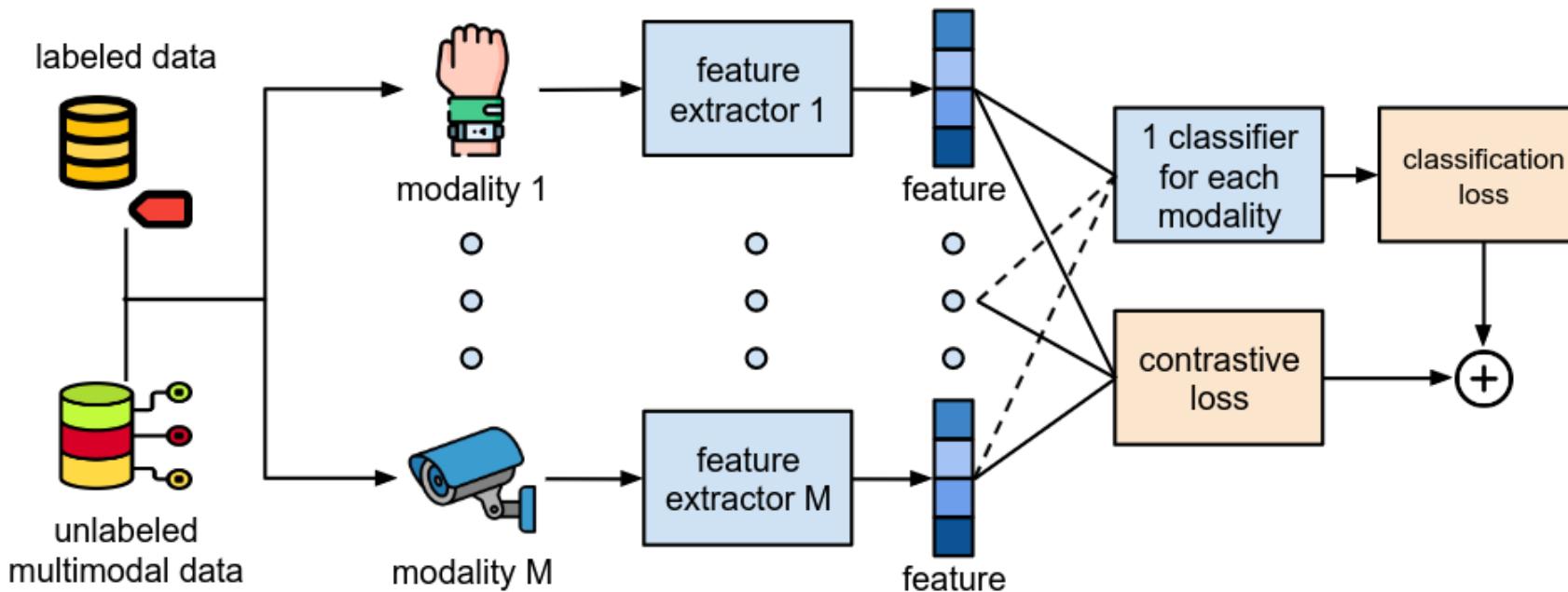


Fig. 1: Overall training process of Virtual Fusion. Dotted lines are optional, depending on label availability.

Virtual Fusion with Contrastive Learning for Single Sensor-based Activity Recognition

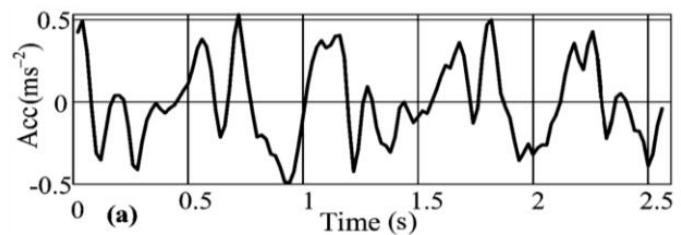
TABLE IV: Comparison on benchmark datasets

Method	UCI-HAR		PAMAP2	
	accuracy	macro F1	accuracy	macro F1
Real time CNN [34]	0.9763	0.9762		
Layer-wise CNN [35]	0.9698	0.9697		
DanHAR [36]			0.9316	
Ensem-HAR [37]	0.9505			
CNN and AOA [38]	0.9523	0.9533		
Marine predators [39]				0.9276
Multi-ResAtt [40]			0.9319	0.9296
DCapsNet [41]	<u>0.9843</u>			
Contrastive Distillation [42]	0.9657	0.9656		
Contrastive Supervision [43]			0.9322	0.9297
AFVF	0.9861	0.9865	0.9672	0.9665

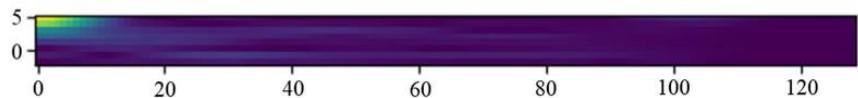
The best and second-best F1-scores are highlighted in bold text and underline respectively.



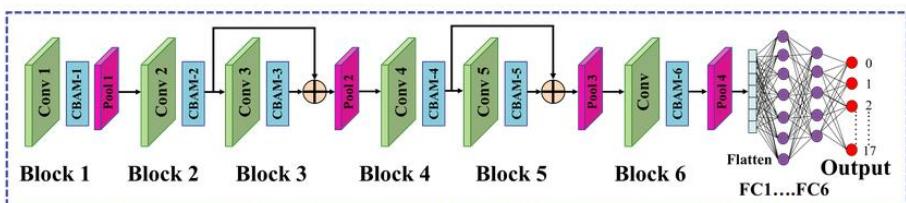
Raw Signal



Processed Signal



AM_DLFC Model Architecture



[Link artigo](#)
[codigo](#)

Table 4. AM-DLFC model performance comparison with existing models using KU-HAR dataset.

KU-HAR						
Research	Accuracy (%)			F1 Score (%)		
[19]	89.67			87.59		
[36]	96.67 (Peak)			96.41 (Peak)		
[34]	88.53			88.52		
[35]	Not Provided			94.25		
Proposed	96.86			96.92		

1	430	0	1	0	0	0
2	2	382	2	0	0	0
3	9	9	333	0	0	0
4	0	0	0	407	18	19
5	0	0	0	44	413	20
6	0	0	0	32	12	442

(a)

1	1727	1	24	25	0	0
2	12	1391	18	11	0	0
3	22	14	412	73	2	3
4	17	4	40	373	0	0
5	0	0	0	1	228	12
6	0	0	3	0	0	201

(b)

Figure 9. The confusion matrix for UCI-HAR and WISDM datasets. (a) UCI-HAR dataset confusion matrix; (b) WISDM dataset confusion matrix.

Sites de Interesse

- <https://ailb-web.ing.unimore.it/icpr/paper/518/nn>

- LSTM-CNN Architecture for Human Activity Recognition [link](#)

<https://arxiv.org/pdf/2011.11151.pdf>

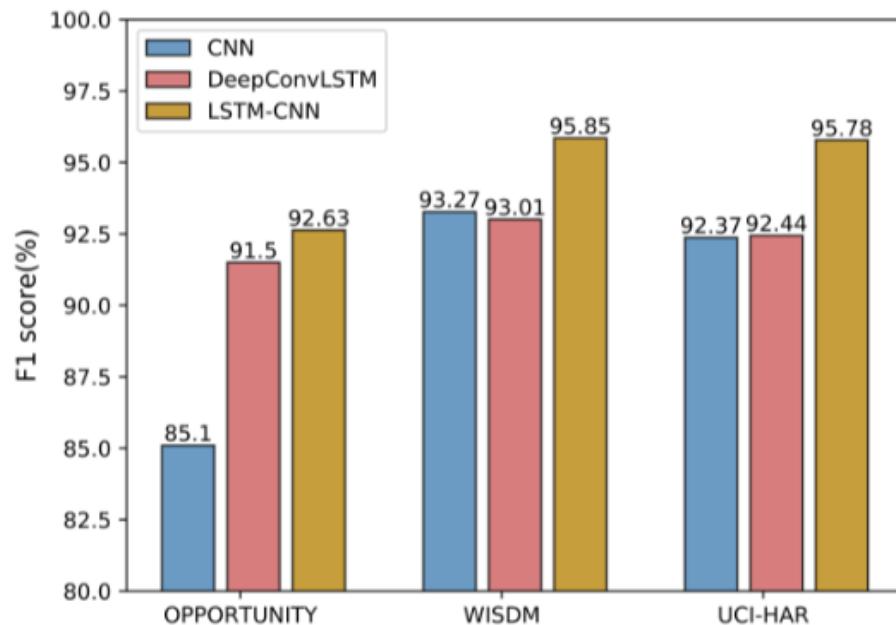
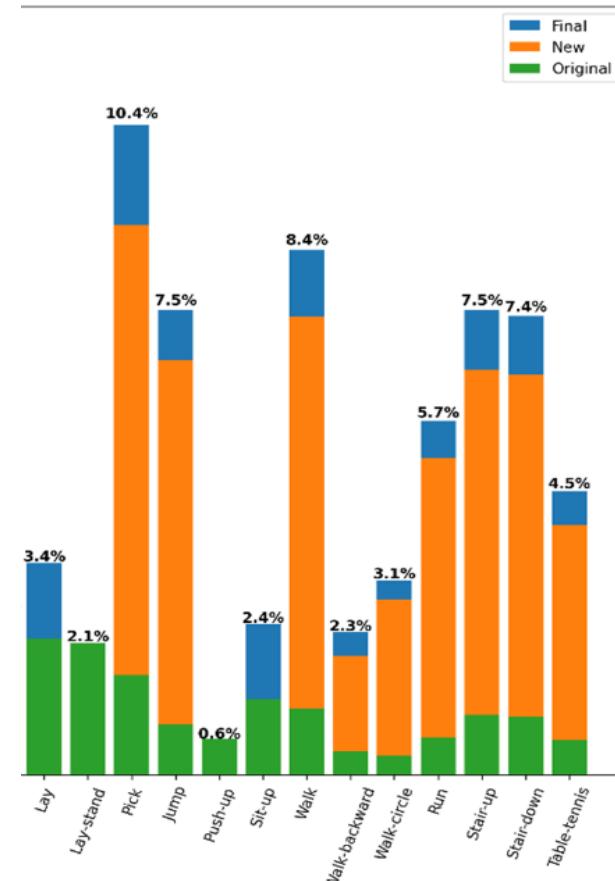


FIGURE 5. Performance of three models on three public datasets.

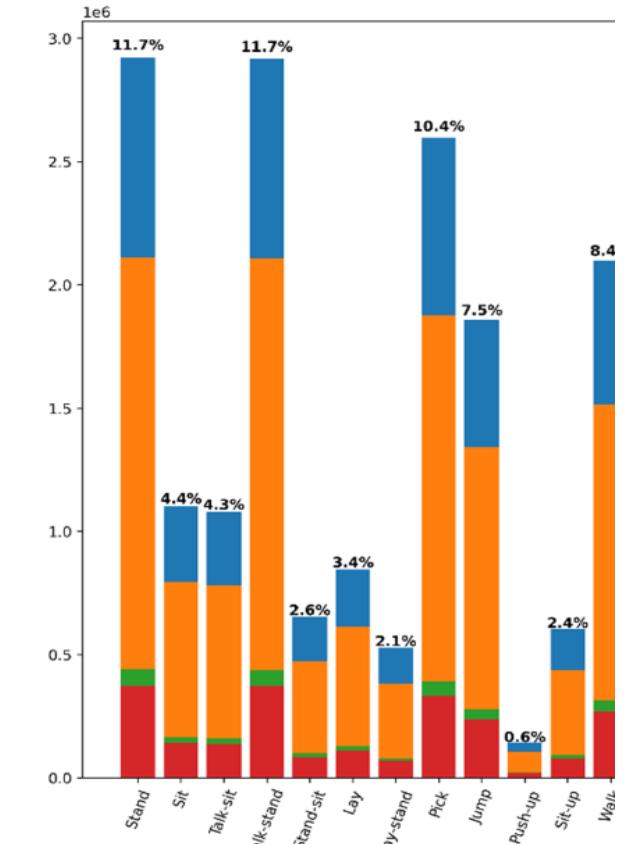
+ · Wearable Sensor-
Based Human
Activity
Recognition with
Transformer Model

[LINK](#)

Wearable Sensor-Based Human Activity Recognition with Transformer Model



(a)

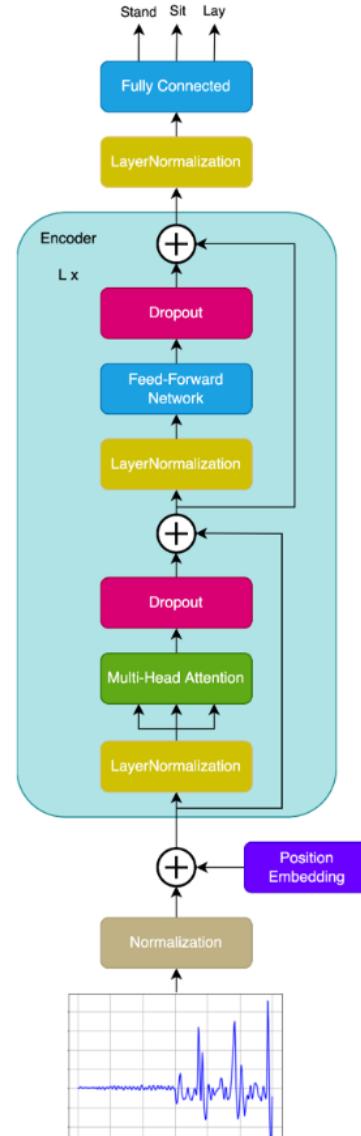


(b)

Figure 2. Distribution of examples by classes: (a) Distribution of examples before and after the data augmentation process; (b) distribution of examples into training, testing, and validation datasets.

Wearable Sensor-Based Human Activity Recognition with Transformer Model

[Codigo Github](#)



Wearable Sensor-Based Human Activity Recognition with Transformer Model

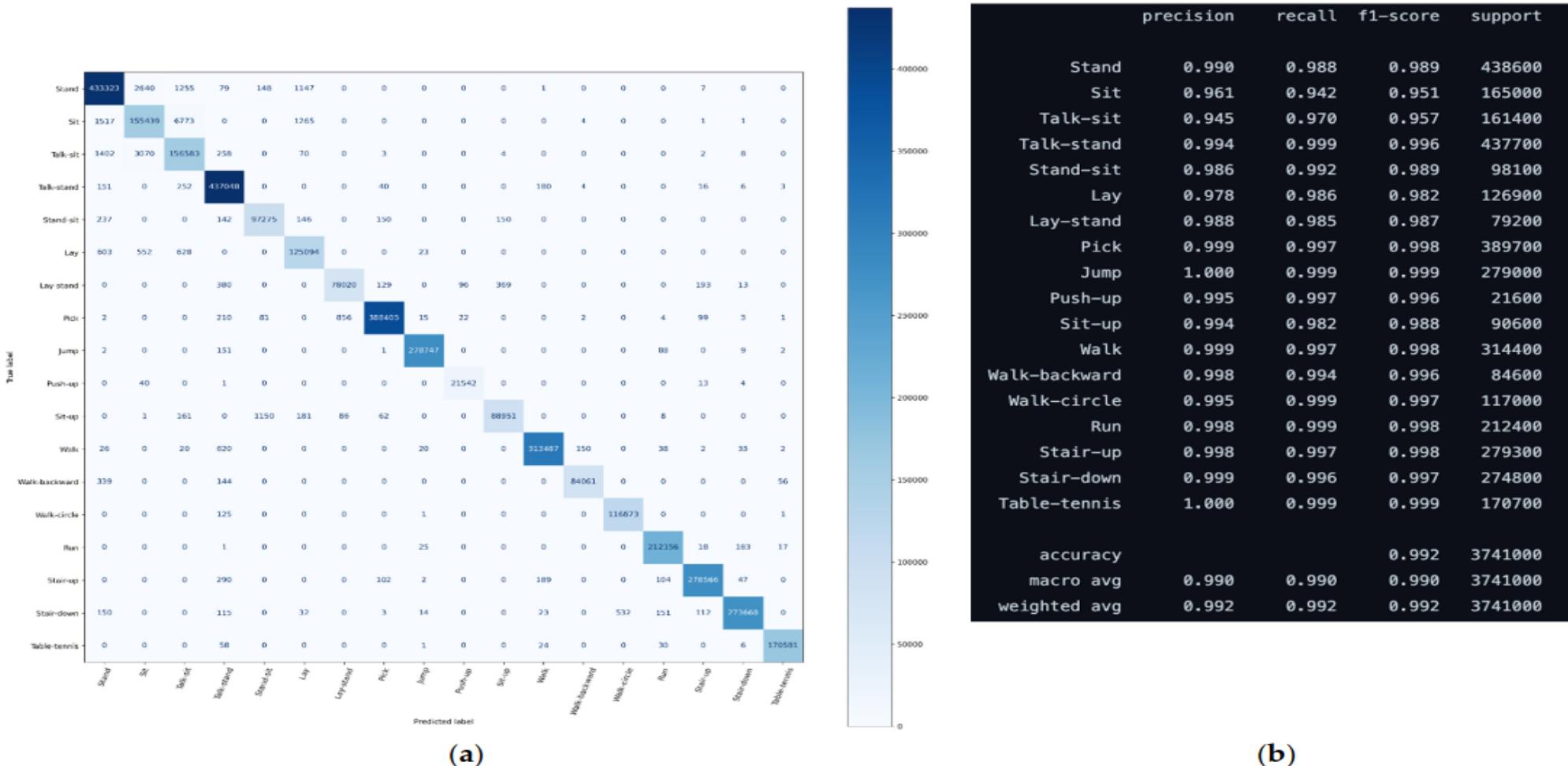


Figure 9. (a) Confusion matrix; (b) class-wise performance of the transformer model for human activity recognition.

The proposed HAR transformer model achieved, on average, 99.2% prediction success as compared with the original 89.67% of the KU-HAR work by [33]. It successfully coped

surveys

- A survey on unsupervised learning for wearable sensor-based activity recognition

Outros artigos

- [Accelerometer based Activity Classification with Variational Inference on Sticky HDP-SLDS](#)
- [DANA: Dimension-Adaptive Neural Architecture for Multivariate Sensor Data](#)
- [Protecting Sensory Data against Sensitive Inferences](#)
- [CROSS-DOMAIN ACTIVITY RECOGNITION VIA SUB-STRUCTURAL OPTIMAL TRANSPORT](#)
- [SenseCollect: We Need Efficient Ways to Collect On-body Sensor-based Human Activity Data!](#)
- [Fine-grained Activity Recognition in Baseball Videos](#)