Neural Networks for Indoor Person Tracking With Infrared Sensors

Neural Networks for Indoor Person Tracking With Infrared Sensors

Osama Bin Tariq,[1*] Mihai T. Lazarescu,[1**] and Luciano Lavagno[1**]

[1] *Department of Electronics and Telecommunications, Politecnico di Torino, I-10129 Torino, Italy*
[*] *Graduate Student Member, IEEE*
[**] *Senior Member, IEEE*

Manuscript received June xx, xxxx; revised June xx, xxxx; accepted July xx, xxxx. Date of publication July xx, xxxx; date of current version July xx, xxxx.

Abstract—Indoor localization has many pervasive applications, like energy management, health monitoring, and security. Tagless localization detects directly the human body, for example via infrared sensing, and is the most amenable to different users and use cases. We evaluate the localization and tracking performance, as well as resource and processing requirements, of various neural network (NN) types. We use directly the data from a low resolution 16-pixel thermopile sensor array in a 3 m x 3 m room, without pre-processing or filtering. We tested several NN architectures, including multilayer perceptron, autoregressive, 1D convolutional neural network (1D-CNN), and long-short term memory. The latter require more resources but can accurately locate and capture best the person movement dynamics, while the 1D-CNN is the best compromise between localization accuracy (9.6 cm root-mean-square error), movement tracking smoothness, and required resources. Hence it would be best suited for embedded applications.

Index Terms—Person localization, tagless localization, thermopile, infrared tracking, CNN, LSTM, autoregressive, multilayer perceptron, embedded neural network, Design Space Exploration.

I. INTRODUCTION

Indoor localization and activity monitoring can be essential for assisted living and domotics, e.g., to increase comfort and reduce energy consumption of appliances, or to check for possibly pathological deviations from daily routines of elderly people. Localization systems that are unobtrusive, privacy-aware, and easy to retrofit can be more easily accepted [1].

Passive infrared (PIR) sensors or thermocouples have been extensively investigated and are often used to detect indoor presence and for localization. PIR sensors are sensitive to movement, while thermopiles can also detect stationary heat sources [2].

Most person indoor tracking approaches with thermopile sensors use machine learning for classification [3]–[5], mathematical modelling [6]–[9], or low resolution image processing [10]–[13]. Classification selects between predefined activities or locations, such as bed exit [3], people count and motion direction [4], or person location on a grid [5].

Tao *et al.* used 43 narrow field-of-view pyroelectric sensors to track daily human activity in a 15 m × 8.5 m space with 0.322 m average error [14]. Kuki *et al.* used 4 pixel × 4 pixel thermopile sensors and fuzzy logic to estimate with 0.246 m average error the walking trajectory of a person in a small 1.58 m × 1.58 m area [15]. Chen *et al.* classified a person location in 60 cm-spaced positions along a snake-like trajectory with 0.134 m mean error, based on the angle of arrival from two 16 pixel × 4 pixel thermopile sensors, processed with multi-frame averaging, background subtraction, and quadratic regression [10]. Shetty *et al.* used 8 pixel × 8 pixel "GridEye" thermopile sensors interpolated to 100 pixels × 100 pixels to demonstrate person tracking with Kalman filters, background subtraction, Gaussian filtering, and iterative thresholds [11]. With the same sensor, interpolated to 71 pixels × 71 pixels and similar processing, Qu *et al.* localized a person with 0.07 m average error while walking on a straight line in a 4 m × 4 m area [12]. Gu *et al.* used a higher resolution 24 pixels × 32 pixels thermopile sensor, interpolated to 93 pixels × 125 pixels, to track with 0.095 m root-mean-square error (RMSE) a person walking on two polygonal paths [13].

We explore the localization and movement tracking accuracy and smoothness for several types of neural networks by drawing inspiration from our earlier exploration of NNs for long-range capacitive sensors for assisted living applications [16], [17]. We search for the architectures and the configurations that perform best while using the least amount of resources, and hence that are the most suitable for embedded processing on low power sensors.

Compared with the previous works, we address the continuous tracking using low-resolution infrared sensors of person movements on extensive arbitrary paths, which resemble closely the walking patterns of a person. Arbitrary trajectory tracking was previously proposed either using tens of sensors [14], or with comparable sensors but relatively low accuracy on small areas [15], or with higher resolution sensors and/or restricted trajectories on comparable areas and with comparable accuracy [10]–[13]. Hence, we provide *either the same accuracy with cheaper sensors, or better accuracy with comparable sensors with respect to the state-of-the-art*. Moreover, we *discuss the resource and performance trade-offs for efficient embedded implementation*.

II. METHODOLOGY, EQUIPMENT, AND TOOLS

We use a 4 pixel × 4 pixel Omron D6T-44L-06 thermopile infrared sensor [18] with temperature resolution 0.06 °C and accuracy ±1.5 °C to monitor the 3 m × 3 m experiment space. It is installed centered on the ceiling, 3.05 m above the floor, having a 2.48 m × 2.57 m field of view (FOV) at floor level.

Similar to our previous work [17], we collected the person reference location with an ultrasound-based tag of the Marvelmind Starter Set HW v4.9 [19], with ±2 cm accuracy, 15 measurements per second. We determined experimentally that the average accuracy of the reference system in our environment is ±3.9 cm (max ±6.4 cm,

Corresponding author: O. Bin Tariq (e-mail: osama.bintariq@polito.it).
M.T. Lazarescu (e-mail: mihai.lazarescu@polito.it) and L. Lavagno (e-mail: luciano.lavagno@polito.it).
Associate Editor:
Digital Object Identifier

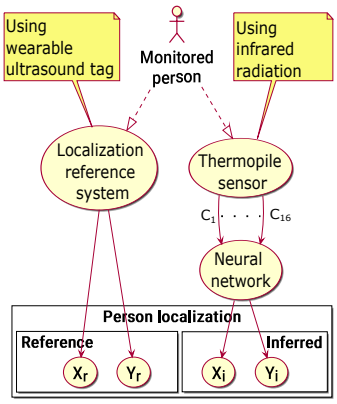


Fig. 1: Experimental data processing uses an accurate ultrasound-based reference (for training data labelling and inference testing), and the thermopile sensor processing with the neural network under test.

standard deviation ±0.7 cm) by measuring the localization accuracy acquiring four times per second for five seconds the location of a person standing on each of 16 equidistant predefined locations inside the 3 m × 3 m experiment room space.

During the actual location tracking experiment, a person walked for 30 min along an arbitrary and fairly irregular path in the space, with variable speed, and we collected synchronous readings at 5 Hz from both the IR sensor and the reference system, as shown in Fig. 1. We collected 9000 tuples, each made of 16 thermal sensor readings and 2 co-ordinates from the reference system relative to the room space.

III. DESIGN SPACE EXPLORATION RESULTS

We test three feedforward NN types: MLP, autoregressive, and 1D-CNN, and one recurrent, LSTM. The neurons use the ReLU activation function, except some LSTM gates that use the default activations [20]. We split the experimental data in 60 % sequential tuples for training, 20 % sequential tuples for validation, 20 % sequential tuples for testing. We used 50 % dropout layers where appropriate to avoid training data overfitting. We train 50 times for each hyperparameter combination with the Keras library and TensorFlow v2.2 back-end, and Adamax first order gradient-based optimization with default parameters. We also tried the Adam optimizer which had an essentially identical RMSE, within ±8 %, or ±0.006 m, of Adamax.

We evaluate (1) the inference quality by the accuracy RMSE, (2) the smoothness of the inferred trajectory by the average of the second derivative [21, p. 62], and (3) the NN computation and memory resource requirements by the total number of operations and parameters. The results are summarized in Table 1 and discussed next.

A. Multilayer Perceptron Neural Networks

Similar to [17], the NN receives one sensor tuple, 16 temperature readings on 16 input neurons, and infers the X and Y co-ordinates of the person in the room on two output neurons. For design space exploration (DSE), we vary the network depth from one to five hidden layers and the number of neurons per hidden layer from 4 to 512, in powers of two.

The best network has three hidden layers with 128 neurons each, inference accuracy RMSE 0.103 m, and smoothness 1.329 m/s^2, which is much higher than the ground truth smoothness (see Table 1).

Table 1: Required parameters (memory), floating-point operations (FLOPs), inference accuracy root mean square error (RMSE) and smoothness for the best configuration of each neural network type

Neural network type	Param.	FLOPs	RMSE (m)	Smooth (m/s^2)
Multilayer perceptron				
128 neurons, 3 layers	35 458	70 149	0.103	1.329
Autoregressive				
1 s win., 64 neur./layer, 3 layers	13 634	26 885	0.117	0.990
1 s win., 256 neur./layer, 3 layers	152 834	304 133	0.098	0.638
1D CNN				
3 s win., 16 filt., 1 conv. lay., ker. 5	4562	8964	0.108	0.305
1 s win., 32 filt., 1 conv. lay., ker. 3	**3810**	**7428**	**0.096**	0.515
LSTM				
3 s window, 2 layers, 64 units	53 890	172 300	0.105	**0.163**
1 s window, 2 layers, 64 units	53 890	172 300	0.109	0.765
Ground truth				
smoothed over 1 s window				*0.443*
smoothed over 3 s window				*0.292*

B. Autoregressive Feedforward Neural Networks

Similar to [17], the NN receives a sliding window of inputs containing multiple sequential 16-sensor reading tuples and infers the X and Y co-ordinates of the middle tuple on two output neurons. The NN accesses both past and future samples, which can help it to learn the movement dynamics. The DSE varies the NN depth from one to five hidden layers, from 4 to 512 neurons per hidden layer in powers of two, and window widths of 1 s and 3 s (covering 5 and 15 tuples respectively, thus changing the input layer size from 80 neurons for the 1 s window to 240 neurons for the 3 s window).

An autoregressive NN with three hidden layers, 256 neurons per layer, and 1 s input window has among the lowest inference RMSE, 0.098 m (see Table 1). But smaller networks, e.g., with 64 neurons per hidden layer, have also small RMSEs, of 0.117 m. Compared to MLP, the autoregressive NN significantly improves the smoothness of the inferred trajectory, to 0.638 m/s^2 from 1.329 m/s^2, thus better capturing the movement dynamics.

C. 1D Convolutional Neural Networks

Convolutional NNs can efficiently extract relevant data patterns and are widely used in image and data time series processing. Efficient pattern recognition helps significantly reduce the computation effort compared to fully connected NNs. Similar to [17], the 1D-CNN receives a sliding window of inputs containing multiple sequential 16-sensor tuples and infers the X and Y co-ordinates of the middle tuple on two output neurons. The NN accesses both past and future samples, which can help it to learn the movement dynamics. The DSE varies the number of kernels from 2 to 64, in powers of two, the kernel size (3, 5, and 7 tuples), the number of convolution layers (1, 2, and 4), and the window width (1 s and 3 s). The hidden layers have convolution layers, an average pooling layer of size five, and a fully connected layer with 64 neurons.

A 1 s window CNN with 32 filters and kernel size of 3 tuples has the best RMSE of 0.096 m and a smoothness of 0.515 m/s^2, both better than the autoregressive NN and requiring only about a quarter of the resources (see Table 1). With a larger 3 s window, the RMSE increases slightly and the smoothness improves markedly, at the expense of more resource requirements.

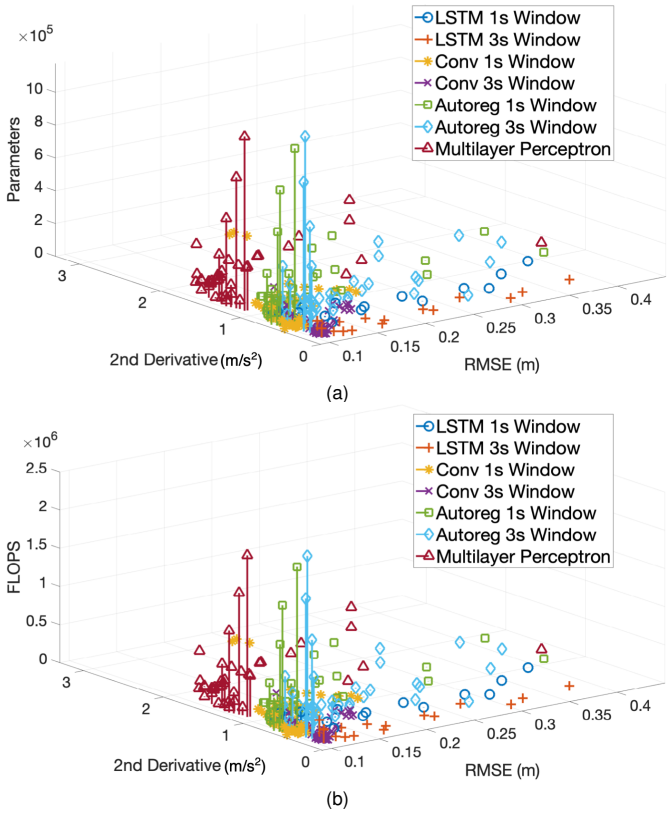


Fig. 2: Inference localization accuracy root mean square error (RMSE) and trajectory smoothness (second derivative) as a function of (a) memory (parameters) and (b) processing (FLOPS) requirements for multilayer perceptron, autoregressive feedforward, 1D convolutional, and long-short term memory (LSTM) neural networks.

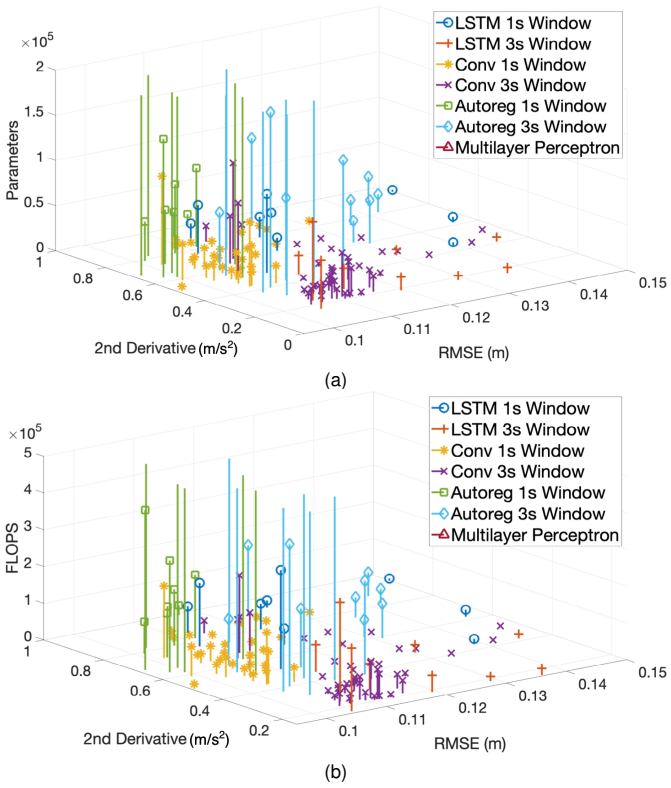


Fig. 3: Detail of inference localization accuracy root mean square error (RMSE) and trajectory smoothness (second derivative) as a function of (a) memory (parameters) and (b) processing (FLOPS) requirements for multilayer perceptron, autoregressive feedforward, 1D convolutional, and long-short term memory (LSTM) neural networks.

D. Long-Short Term Memory Neural Networks

LSTMs are recurrent networks used mostly where history and context awareness can improve the inference, e.g., for handwriting and speech recognition, or translation. Similar to previous work [17], the LSTM receives a sliding window of inputs containing multiple sequential 16-sensor tuples and infers the X and Y co-ordinates of the middle tuple. In the DSE we vary the LSTM layers (1, 2, and 3), LSTM units from 2 to 64, in powers of two, and the input window width (1 s and 3 s).

The LSTM achieves by far the best smoothness, 0.163 m/s^2, with a good RMSE of 0.105 m using a 3 s input window (see Table 1), but requires 15 to 20× more resources than the 1D-CNN. With a smaller window of 1 s, the RMSE changes only slightly to 0.109 m, but the smoothness lowers significantly to 0.765 m/s^2 for virtually the same resource requirements (being a recurrent network, the resource requirements are largely independent on the input window size).

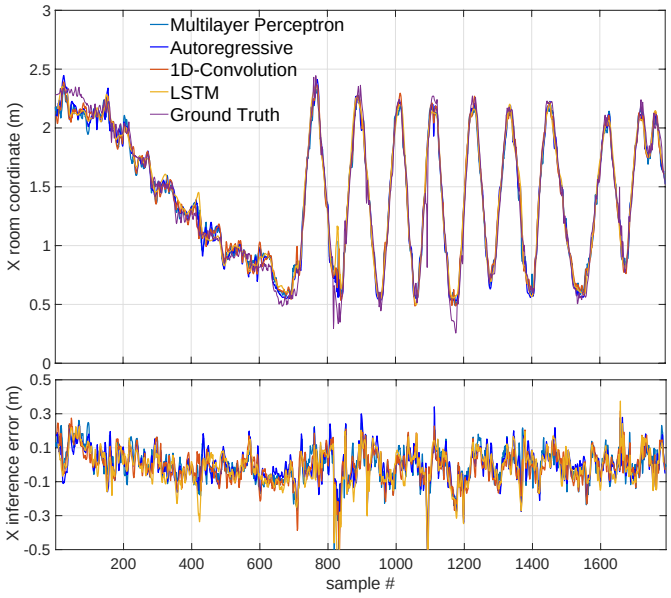
IV. RESULT DISCUSSION AND OPTIMIZATIONS

We summarize in Fig. 2 (further zoomed around the origin in Fig. 3) the dependence of a) memory requirements (parameters) and b) processing requirements (FLOPs) on the localization inference RMSE and the inference smoothness (2nd derivative) for all the NNs. Each NN allows a distinct performance-resource trade-off (note that memory and processing are closely correlated). The MLP NNs can have low RMSEs, but mostly poor smoothness and high resource requirements. The autoregressive NNs have better inference smoothness, especially with 3 s windows, but still high resource requirements (see Fig. 3). The 1D-CNN and LSTM NNs perform best. The former generally have better performance-resource trade-offs with 3 s windows.

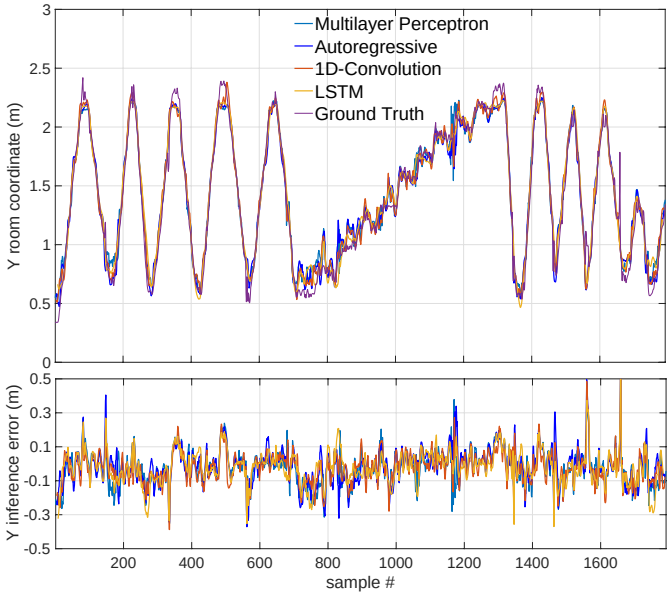
Fig. 4 comparatively shows the inference of the X and Y co-ordinates of the person while moving within the space. The MLP and the autoregressive NNs seem to be the most "noisy." The LSTM looks the smoothest, but leaves some extremes uncovered, while the 1D-CNN seems a good compromise between trajectory coverage and smoothness.

Considering the above, the 1D-CNN with 3 s input window seems the best trade-off between inference performance and resource requirements for embedded implementation (see Table 1). Moreover, our most accurate tracking inference over an area of 3 m × 3 m using one 4 pixel × 4 pixel sensor has 0.096 m RMSE (see Table 1), which is sufficient for most assisted living or home automation applications.

Comparatively, the reports in the state-of-the-art can use tens of sensors to monitor a space 15 times larger with a much higher average error of 0.322 m [14], or use two higher-resolution sensors to classify the location in predefined 60 cm-spaced positions with higher mean error of 0.134 m [10], or a sensor with comparable resolution over a quarter of our monitored space, with much higher average error of 0.246 m [15], or more expensive sensors with four times higher resolution, further enhanced with interpolation, achieving comparable localization accuracy over comparable areas, but for predefined (not arbitrary) trajectories [11], [12].



(a)



(b)

Fig. 4: Ground truth, trajectory tracking inference and its error for the (a) X axis and (b) Y axis for the best NN configurations

V. CONCLUSION

Low resolution infrared camera sensors can be used for low cost privacy-aware indoor person localization and movement tracking using neural networks. Network architecture and hyperparameter values greatly influence the sensor performance. This paper explores trade-offs between location accuracy, trajectory smoothness, computing cost and memory resources, in order to find the best compromise for embedded implementations with limited resources.

The networks that consider a sequence of sensor readings, such as the autoregressive, 1D-CNN, or LSTM, have smoother inferences that better follow the actual dynamics of the movements of a person. Among these, the recurrent networks, such as LSTMs, can consider a longer movement history and achieve the best inference smoothness, 0.163 m/s^2, and 0.105 m localization accuracy RMSE. However, the 1D-CNN with a 1 s input window has the best localization accuracy, of 0.096 m RMSE, needs much fewer computing resources (7428 FLOPs compared to 172 300 FLOPs for LSTMs) and memory resources (3810 parameters compared to 53 890 parameters for LSTMs), thus being better suited for embedded implementation.

In future work, movement tracking can be extended to multiple persons using NNs such as Yolo for detection, and machine learning algorithms such as Support Vector Machines for movement tracking.

REFERENCES

[1] F. Alam, N. Faulkner, and B. Parr, "Device free localization: A review of non-rf techniques for unobtrusive indoor positioning," *IEEE Internet of Things Journal*, 2020.

[2] D. Xu, Y. Wang, B. Xiong, and T. Li, "Mems-based thermoelectric infrared sensors: A review," *Frontiers of Mechanical Engineering*, vol. 12, no. 4, pp. 557–566, 2017.

[3] S. Chiu, J. Hsieh, C. Hsu, and C. Chiu, "A convolutional neural networks approach with infrared array sensor for bed-exit detection," in *2018 International Conference on System Science and Engineering (ICSSE)*, 2018, pp. 1–6.

[4] C. Basu and A. Rowe, "Tracking motion and proxemics using thermal-sensor array," *arXiv preprint arXiv:1511.08166*, 2015.

[5] C. Kowalski, K. Blohm, S. Weiss, M. Pfingsthorn, P. Gliesche, and A. Hein, "Multi low-resolution infrared sensor setup for privacy-preserving unobtrusive indoor localization." in *ICT4AWE*, 2019, pp. 183–188.

[6] J. Kemper and D. Hauschildt, "Passive infrared localization with a probability hypothesis density filter," in *2010 7th Workshop on Positioning, Navigation and Communication*. IEEE, 2010, pp. 68–76.

[7] H. M. Ng, "Human localization and activity detection using thermopile sensors," in *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2013, pp. 337–338.

[8] X. Zhang, H. Seki, and M. Hikizu, "Detection of human position and motion by thermopile infrared sensor," *International Journal of Automation Technology*, vol. 9, no. 5, pp. 580–587, 2015.

[9] L. Wu and Y. Wang, "Compressive sensing based indoor occupancy positioning using a single thermopile point detector with a coded binary mask," *IEEE Sensors Letters*, vol. 3, no. 12, pp. 1–4, 2019.

[10] W.-H. Chen and H.-P. Ma, "A fall detection system based on infrared array sensors with tracking capability for the elderly at home," in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 2015, pp. 428–434.

[11] A. D. Shetty, B. Shubha, K. Suryanarayana *et al.*, "Detection and tracking of a human using the infrared thermopile array sensor—"Grid-EYE"," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. IEEE, 2017, pp. 1490–1495.

[12] D. Qu, B. Yang, and N. Gu, "Indoor multiple human targets localization and tracking using thermopile sensor," *Infrared Physics & Technology*, vol. 97, pp. 349–359, 2019.

[13] N. Gu, B. Yang, and T. Li, "High-resolution Thermopile Array Sensor-based System for Human Detection and Tracking in Indoor Environment," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1926–1931.

[14] S. Tao, M. Kudo, H. Nonaka, and J. Toyama, "Recording the activities of daily living based on person localization using an infrared ceiling sensor network," in *2011 IEEE International Conference on Granular Computing*. IEEE, 2011, pp. 647–652.

[15] M. Kuki, H. Nakajima, N. Tsuchiya, and Y. Hata, "Human movement trajectory recording for home alone by thermopile array sensor," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2012, pp. 2042–2047.

[16] O. B. Tariq, M. T. Lazarescu, and L. Lavagno, "Neural network-based indoor tag-less localization using capacitive sensors," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 9–12.

[17] O. Bin Tariq, M. T. Lazarescu, and L. Lavagno, "Neural networks for indoor human activity reconstructions," *IEEE Sensors Journal*, 2020.

[18] "D6t mems thermal sensors." [Online]. Available: https://www.components.omron.com/product-detail?partNumber=D6T

[19] "Starter Set HW v4.9." [Online]. Available: https://marvelmind.com/product/starter-set-hw-v4-9/

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] J. Ramsay, G. Hooker, and S. Graves, *Functional data analysis with R and MATLAB*. New York, NY: Springer, 2009.