POLITECNICO MILANO 1863
SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Anomaly Detection in Elderly Behavior Using TinyML and UWB Radar

Laurea Magistrale in Computer Engineering - Ingegneria Informatica

Author: Alessandra Fiore
Advisor: Prof. Manuel Roveri
Co-advisor: Massimo Pavan
Academic year: 2022-2023

---

1. Introduction

Machine learning (ML) plays a crucial role in developing intelligent and inventive solutions across diverse sectors. The realm of TinyML is empowering the deployment of such cutting-edge technologies onto resource-constrained devices.

In the context of elderly care, there is an emerging need for advanced technologies that can improve health care delivery. ML technologies can provide valuable support for health carers in identifying significant changes in the conditions of patients under their care.

The primary goal of this thesis is to create an effective solution for monitoring the behavior of older individuals in order to detect behavioral anomalies early and deliver relevant notifications. To achieve this goal, we designed a novel smart sensor solution, encompassing an Ultra-Wide-Band (UWB) radar sensor and an *in-sensor tinyML algorithm*, and an *anomaly detection algorithm* that uses the outputs of sensors in the house.

The *TinyML algorithm*, exploiting the UWB-radar data, allows us to assess both the presence and distance of an individual within the various rooms, thus enabling comprehensive and detailed monitoring of movement and activity within the home environment. Successively, we propose the usage of an *anomaly detection algorithm* on an edge device that can identify deviations from the standard habits of the patient by collecting data in output from the TinyML algorithm through the first period of use of the device, personalizing it on the habits and schedules of the specific patient.

After this phase, the algorithm becomes able to notify a carer about possible anomalies occuring in the house of the elder.

2. Backgroung

2.1. TinyML

Tiny machine learning (TinyML) is an innovative domain in machine learning. It entails condensing complicated deep learning models to fit on IoT devices and microcontrollers (MCUs), extending the scope of AI applications. This approach offers several significant advantages. It facilitates machine learning with a minimal memory foot-print, often utilizing only a few hundred kilobytes. By employing TinyML for ultra-low power tasks at the edge, a new approach called ML sensor emerges, providing a design where AI functionalities are not simply appended via cloud or mobile connections but are integrated within the sensor device itself. ML sensors provide a new paradigm for sensing by moving processing and analysis to the device rather than the cloud. This strategy gives priority to the proximity of data resulting in lower latency and increased data privacy as raw data is never transmitted.

Despite the benefits, there are some challenges to navigate: the efficiency of deep learning models often comes with significant computational requirements, which prevent their integration into TinyML applications due to the severe resource limitations of devices like microcontrollers [4, 6].

These challenges can be efficiently addressed by employing neural network compression methods, such as quantization and pruning. Quantization minimizes the network's memory footprint by decreasing the number of bits needed to represent neuron weights, whereas pruning removes less consequential neural connections. This process reduces the model's complexity without significantly affecting its performance.

2.2. UWB radar

Ultra-Wideband (UWB) radar belongs to a class of radars that uses radio systems characterized by very wide bandwidths. This means that it enables the frequency spectrum to be shared among different users, helping to manage interference and signal path complexity in radio communications [5].

The unique ability of UWB radar to pass through obstacles makes UWB as a promising technology for ensuring accurate localization, particularly in cluttered and challenging scenarios. It stands out for its exceptional accuracy, which allows it to precisely locate individuals and objects within a range of just a few centimeters [1].

3. Related Literature

In the scientific literature, the problem of indoor location has received considerable attention, highlighting the importance of developing reliable indoor positioning system. Some methods consider the use of wearable sensors by the subject or the use of cameras to monitor the surroundings. For example, Soo-Cheol Kim et al. introduce a localization method based on radiofrequency identification (RFID) technology to accurately monitor the elderly. The RFID system consists of readers and tags. Precisely an RFID tag affixed to living beings comprises an antenna designed to both receive and transmit RF signals, while an RFID reader establishes communication with one or multiple tags within its range, transmitting the acquired data to the backend server for additional processing [2].

The use of UWB radar for indoor tracking, as in this study, represents a totally non-intrusive system, eliminating the need for wearable tags or sensors. This feature gives the system a significant advantage in terms of convenience and comfort for users, as it does not require any additional device to be worn. A significant contribution in this area is the study conducted by Samuel G. Leitch et al. who have shown that among the Wi-Fi, Bluetooth Low Energy (BLE), Inertial Measurement Unit (IMU) and UWB technologies, only WiFi and UWB demonstrate the ability to achieve an average estimation error of less than 10 cm [3].

While these studies have investigated the use of UWB technology for indoor localization, none have designed and developed a solution specifically for tiny devices. Such a solution is able to perform operations or calculations quickly and efficiently using a limited amount of computational resources.

Although numerous studies have explored the application of UWB technology for indoor localization, as highlighted by the cited study, and there are several solutions that take advantage of UWB technologies integrated into resource-constrained devices, the presented distributed solution encompassing smart, tinyML enabled, UWB-radar sensors present a promising and unique approach to the enhancement of elderly care, offering a privacy-preserving and non-invasive solution that can enhance the quality of life for seniors while at the same time limiting the burden on their carers.

4. Problem Formulation

The main objective of this research is to develop an advanced system for monitoring the behavior of elderly people within their homes so that any anomalies or deviations from normal behavior can be quickly identified. The problem concerns

the analysis of data acquired from a UWB radar, particularly, the most recent k radar acquisitions, being k, a parameter specific to the application, that is chosen by the designer.
The problem can be reformulated as the design of a system $f_v(x_t, x_{t-1}, x_{t-2}...x_{t-(k-1)})$ able to map the frames of the radar $(x_t, x_{t-1}, x_{t-2}...x_{t-(k-1)})$ into the label A_t, being k defined as the length of the observation window, x_t the radar acquisition at time t with dimensions $N \times M$, where N denotes the number of pulses produced and, therefore, returned to the radar, and M denotes the number of bins, or discrete intervals of space, and A_t the label where $A_t \in \{abnormal, normal\}$. Abnormal in this context refers to a pattern of behavior that deviates significantly from the data collected in the first period of device use, which is aimed at collecting the subject's habits.

5. Proposed Solution

The problem has been divided into two subproblems: the design of a sensor based on UWB and tinyML, which by analyzing data from the UWB radar is able to recognize the presence of a subject in the room and its distance from the radar, and an algorithm that performs anomalous pattern detection in time series composed of the outputs of these sensors deployed in the various rooms.
The proposed solution for the sensor problem involves the use of two TinyML algorithms, that are deployed on the embedded devices placed in each room and analyze the radar acquisitions:

- The *in-sensor presence detection algorithm*, that aims to evaluate the presence or absence of the subject.
- The *in-sensor distance estimation algorithm*, which provides detailed information on its position in the room in the case of detected presence.

Subsequently, an edge device aggregates the results generated from the TinyML algorithms to compute information related to the collected data, such as the mean and standard deviation of distances recorded for each room in a given time window. These new insights are then utilized to construct a time series, which will be fed into the multi-device *anomaly detection algorithm*. This algorithm will evaluate the time series to determine whether it deviates from the subject's typical behavior, consequently classifying it as anomaly or normal.

5.1. In-sensor Presence Detection

Let $s_t \in \mathbb{R}^{N \times M}$ denote the acquisition obtained by the UWB radar, where $N, M \in \mathbb{N}$ represent the dimensions, where N denotes the number of pulses produced and, therefore, returned to the radar, and M denotes the number of bins, or discrete intervals of space.
The problem is assigning to data s_t the label y_t, where $y_t \in \{absent, present\}$.

$$y_t = P(s_t) = \begin{cases} 0 & absent \\ 1 & present \end{cases}$$

The dataset obtained through UWB radar acquisitions consists of a total of 805 data, each lasting 1.1 seconds. Of these, 322 cases corresponding to "non-presence" situations, in which there were no objects detected within the radar range, and remaining number of data, 483, represents "presence" situations, in which a subject was positioned within the range of the radar. This ensure a balanced representation between the two classes.
Before feeding the data to the *in-sensor presence detection algorithm* it is essential to perform a preprocessing step on the it in order to make it informative and clean. Data point acquired by the radar UWB, which is rich in useful information, has a size of 11x248, where 248 denotes the space dimension (spatial bins) and 11 the time dimension (1.1 seconds). The spatial dimension contains a representation of signal amplitude and phase alternated, and thus can be seen as 124 complex values. Then absolute value was computed on the entire matrix to refine the representation by preserving only the amplitude information of the signal. Consequently, the matrix was reduced to a size of 11x124 real values. Another crucial phase in processing the data involved employing decluttering process, an essential technique designed to eliminate noise from the dataset, preserving only pertinent information. The employed declutter method is the "moving average filter" technique. After the decluttering phase, each data matrix was subjected to targeted trimming to exclude the least informative parts of the record, that didn't provide information regarding the subject's behaviors, such as those contained in the

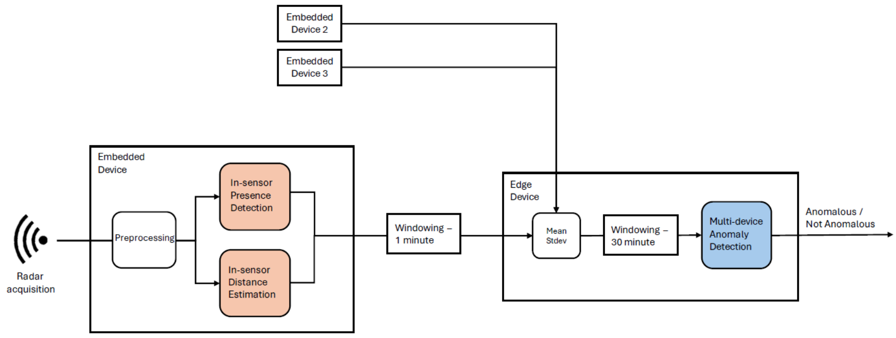

Figure 1: Proposed solution.

near and far spatial bins, so, a matrix of size 11x56 was generated. In the last data preprocessing step, each sample undergoes normalization before being used as input for the algorithm. Normalization is the process of changing the values of a data set so that they have a common scale in order to facilitate data comparison and analysis.
At the conclusion of this preprocessing step, the data are ready to be used by the *in-sensor presence detection algorithm.*
To handle the task of distinguishing between presence and absence, an approach involving the use of a Convolutional Neural Network (CNN) functioning as a binary classifier was implemented. Its input layer is a Conv2D layer specifically engineered to process input images formatted as two-dimensional matrices sized 11x56. It used 16 3x3 filters and relu activation function. Then, a second Conv2D, now with 32 filters, is employed for detecting more complex and abstract patterns. The class head of the neural network consists of a single dense layer with one node and sigmoid activation function. This layer produces a number between 0 and 1, with 1 denoting a high probability of presence and 0 a low probability.

5.2. In-sensor Distance Estimation

This problem uses the same UWB radar acquisition $s_t \in \mathbb{R}^{N\times M}$ previously described for the presence detection problem. In particular, data acquisitions were performed at three target distances: 0.50 meters, 1.25 meters and 2.50 meters. In addition, also small angular ranges were considered. Specifically, from -30 to 30 degrees with respect to the axis of the radar at 1.25 meters and -15 to 15 degrees at 2.5 meters.

This time, the goal is to assign to the data s_t a real value d representing the distance between the subject and the radar.

$$d = D(s_t)$$

The data preprocessing step is also identical to that described for the presence detection problem.
A second Convolutional Neural Network was employed to solve the distance estimation problem. It has the same two Conv2D layers, but in contrast to the network utilized for presence/absence classification, this particular network was set up for regression. As output layer it uses a single neuron with liner activation function that produces a continuous value as output. This implies that the objective of the neural network is to forecast a continuous variable (distance) rather than a discrete one.
Finally, a total of 1725 data points have been recorded during data collection exploited for the CNN. These data were distributed among three distances, with 437 collected positioning the target at 0.50 meters, 644 at 1.25 meters, and a further 644 at 2.50 meters. This method tries to enable the convolutional neural network to effectively estimate continuous distances within the certain range.

5.3. Anomaly Detection for Time Series

The problem aims to apply anomaly detection techniques in a multivariate time series T_i, which describes the behavior of the subject in an example apartment and in a specific time window. The time series T_i contains the key information presence, mean distance, and standard deviation for each room and for each time step, calculated

from the collection of the outputs of the tinyML algorithms.
The *multi-device anomaly detection algorithm* assigning the value A_t to T_i, where A_t is the output of the algorithm and $A_t \in \{abnormal, normal\}$.

$$A_t = AD(T_i) = \begin{cases} 0 & abnormal \\ 1 & normal \end{cases}$$

Using the outputs obtained with the two tiny NNs, we compute the mean and standard deviation of the distances within each interval of granularity of 1 minute only for data in which the presence was detected. Utilizing the mean of distances provides an approximation of the subject's average positioning within each time interval, while the standard deviation offers an indication of the variability of these positions and the amount of movement performed by the target. Following this, we proceeded to structure this information into temporal windows, composed of the data collected during each granularity interval. These time windows function as multivariate time series, enabling for a more detailed analysis of behaviors over time. By examinating these time series, specific patterns can be identified and the presence of anomalous behaviors can be assessed. In particular, the time series are analyzed in windows of 30 minutes. Before the time series is given as input to the *anomaly detection algorithm*, for each granularity interval in the series, the corresponding time is calculated in trigonometric format, that is, expressed in hours and minutes in the form of sine and cosine. This continuous representation of time highlights the cyclic nature of time, preserving temporal information and allowing the algorithm a more effective interpretation of time.
The anomaly detection algorithm consist of a One Class SVM used to identify unexpected behaviors.
The dataset used for training the OCSVM model was generated by exploting the outputs of the in-sensor presence and distance estimation algorithms, simulating their usage on a real-world apartment composed of 3 rooms. This dataset consists of 30-minute time series, each represented by a 13x30 matrix. Of the 13 rows of the matrix, the first 9 rows represent the presence, mean, and standard deviation information for the 3 rooms; the last 4 rows are used for the trigonometric representation of time Each column of the matrix corresponds to a single minute summarizing values of the sensors (presence, average distance, standard deviation). Finally, the dataset used for the OCSVM has 2160 data in total composed mostly of normal data and several anomalies used to test the model's ability to identify outliers.

6. Results

The results presented show the performance metrics attained by the described models.

6.1. Results of Presence Detection

Table 1 demonstrates the accuracy obtained by the CNN employed in the presence detection problem across the three different sets (training, validation, test).

train acc	val acc	test acc
0.9475	0.9304	0.8768

Table 1: CNN classifier results.

After training, the model undergoes a quantization process to reduce its memory consumption. Specifically, quantization is applied to the model's weights, converting the values from floating point to integers. This results in a reduction in the size of the model: from a memory occupancy of 136360 B to only 28936 B without any loss of accuracy in the model. This claim was verified by running inference on the same test set after conversion, showing that the quantized model successfully retains its predictive ability.
A simple method was tried for the purpose of classification before using the machine learning model. This method calculated the peak value for each data in the dataset and compared this maximum with a threshold (maximum peak value among all data representing the "non-presence" in the dataset). If the calculated peak exceeded the threshold, the data was classified as "presence," otherwise "non-presence". However, this method achieved a lower accuracy than the adopted ML model, with an accuracy value obtained on the test set of 0.7609, highlighting how such a simple approach is unable to understand the complexity of the data.

6.2. Results of Distance Estimation

The CNN used for solving the distance estimation problem, is evaluated based on Mean Absolute Error (MAE), a metric used to evaluate the accuracy of predictions in regression problems.

train mae	val mae	test mae	quantized mae
0.1243	0.1199	0.1451	0.8769

Table 2: CNN regression results.

This model underwent identical quantization as the CNN model employed for presence detection. Subsequent to quantization, the memory footprint was substantially reduced from 89800 B to 16992 B. Nevertheless, this led to a degradation in the model's accuracy, as seen in Table 2 showing the increase in mae on the test set.
A simple approach was also tried for the distance estimation problem before using the ML algorithm. This method aimed to detect the bin where the data had the peak value and, based on the peak bin, to estimate the discrete position of the subject (near, middle, far). Although the level of accuracy obtained is good (0.9368 on the test set), this approach was not able to provide a continuous distance value, but rather only provides a discrete label.

6.3. Results of Anomaly Detection

The OCSVM has demonstrated strong competence in identifying outlier data, confirming the effectiveness of the methodology used, with an accuracy on the test set of 98.06%. The test set consists of 103 data including normal and anomalous data to test the model's ability to identify as anomalous the data that deviate from those used for training.
Another method tested for the purpose of anomaly detection was the use of three different SVM models each aimed at identifying a specific type of anomaly (SVM_{double} for the anomaly of the double appearances in two different rooms, SVM_{abs} for the anomaly of the absence at unscheduled times and SVM_{diff} for the anomaly of the presence in a room other than the scheduled room). However, the use of different SVMs requires a priopri knowledge of anomalous behavior, thus failing to generalize should a new, nonspecific anomaly arise.

Specific datasets were also created for the three SVM models. In particular, each dataset used for the SVMs contains normal data and anomalous data representing the specific anomaly for which we are training the SVM. The dataset used for the SVM_{diff} consists of 1170 data points in total, the dataset for the SVM_{abs} includes 1472 data points, while the dataset for the SVM_{double} includes 1148 data.
The SVMs used for detecting specific anomalies have also shown high rates of accuracy. Each of these SVMs is tasked with addressing specific anomalies, resulting in simpler, more specialized models that contribute to higher performance in anomaly detection.
Nevertheless, when tested on a new type of anomalous data representing presence at unscheduled times, the three SVMs, which were designed to detect other types of anomalies, show significat lower results w.r.t the OCSVM . This result showed the inability of the SVM to generalize on other type of data, due to their specific design to detect specific types of anomalies. In contrast, the OCSVM showed superior capabilities in detecting the presence of unusual anomalies, with significantly better results than the other three SVMs. Table 3 shows and compares all the results obtained by SVMs and OCSVM.

	Specific SVMs	OCSVM
Double Presence	0.9539	1.0
Unexpected Absence	1.0	0.9333
Different Room	0.9933	1.0
Average	98.24	97.78
New Problem	0.5229*	1.0

Table 3: SVMs and OCSVM results.

* Average of the 3 SVMs.

7. Conclusion

In conclusion, we have seen how Tinyml combined with UWB technology leads to an excellent aftermath in the field such as presence detection and indoor localization, enabling accurate and noninvasive monitoring, and how this combination is therefore suitable in the world of healthcare and indoor monitoring.
Of particular importance to this thesis is the use of OCSVM for anomaly detection. It plays a key role in detecting abnormal behavior patterns. Its ability to distinguish normal behavior

from potential risks greatly increases the effectiveness of our system in keeping the elderly safe. In short, we developed a distributed solution that can combine the use of UWB and TinyML for patient monitoring and an anomaly detection algorithm to detect deviations in patient behavior. In the future, we expect that as these technologies advance, these solutions will be increasingly used in the elderly care context to meet the growing demand of an aging population.

References

[1] Davide Dardari, Chia-Chin Chong, and Moe Win. Threshold-based time-of-arrival estimators in uwb dense multipath channels. *IEEE Transactions on Communications*, 56(8):1366–1378, 2008.

[2] Soo-Cheol Kim, Young-Sik Jeong, and Sang-Oh Park. Rfid-based indoor location tracking to ensure the safety of the elderly in smart home environments. *Personal and Ubiquitous Computing*, 17(8):1699–1707, Dec 2013.

[3] Samuel G. Leitch, Qasim Zeeshan Ahmed, Waqas Bin Abbas, Maryam Hafeez, Pavlos I. Laziridis, Pradorn Sureephong, and Temitope Alade. On indoor localization using wifi, ble, uwb, and imu technologies. *Sensors*, 23(20), 2023.

[4] Ji Lin, Ligeng Zhu, Wei-Ming Chen, Wei-Chen Wang, and Song Han. Tiny machine learning: Progress and futures [feature]. *IEEE Circuits and Systems Magazine*, 23(3):8–34, 2023.

[5] K. Siwiak. Ultra-wide band radio: introducing a new technology. In *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*, volume 2, pages 1088–1093, 2001.

[6] Matthew Stewart, Pete Warden, Yasmine Omri, Shvetank Prakash, Joao Santos, Shawn Hymel, Benjamin Brown, Jim MacArthur, Nat Jeffries, Sachin Katti, Brian Plancher, and Vijay Janapa Reddi. Datasheets for machine learning sensors: Towards transparency, auditability, and responsibility for intelligent sensing, 2024.