Fine-grained Human Activity Recognition through Dead-reckoning and Temporal Convolutional Networks

Nicolò La Porta*
nicolo.laporta@supci.ch
Faculty of Informatics, Università della Svizzera Italiana
Lugano, Ticino, Switzerland
Institute of Information Systems and Networking, SUPSI
Lugano, Ticino, Switzerland

Luca Minardi
luca.minardi@supsi.ch
Institute of Information Systems and Networking, SUPSI
Lugano, Ticino, Switzerland

Michela Papandrea
michela.papandrea@supsi.ch
Institute of Information Systems and Networking, SUPSI
Lugano, Ticino, Switzerland

ABSTRACT

Human Activity Recognition (HAR) represents an important task for many healthcare applications. From the perspective of developing patient-specific solutions, it is clear how the use of artificial intelligence enhances the potential of HAR. The present work settles its roots in the context of early-diagnosis of neurodevelopmental disorders in children (Autism Spectrum Disorder, ASD) and in the evaluation of their motor skills. In this paper, we present an artificial intelligence-based approach for fine-grained HAR which relies on dead-reckoning applied to data collected through inertial measurement units (IMUs). This approach has been applied on a dataset collected through IMU-embedded toys in order to validate its feasibility in the inference of infants fine-grained motor skills. The proposed solution's workflow starts from the estimation of the orientation and position of solid objects through dead- reckoning exploiting Kalman filters and moves to the extraction of informative features, which are then used to feed a Temporal Convolutional Network (TCN). The achieved training average accuracy of 89% highlights how such a non-intrusive approach reaches great performances on HAR tasks, even overcoming the limitations of most of the works already present in literature, based on wearable sensors and/or computer vision techniques. The presented work and achieved results represent a solid base for IoT-based systems aiming at supporting clinicians in the early diagnosis of ASD in children.

KEYWORDS

Human Activity Recognition, Dead-Reckoning, IMU, Deep Learning, Temporal Convolutional Networks

ACM Reference Format:
Nicolò La Porta, Luca Minardi, and Michela Papandrea. 2024. Fine-grained Human Activity Recognition through Dead-reckoning and Temporal Convolutional Networks. In *iWOAR 2024 - 9th international Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR 2024), September 26–27, 2024, Potsdam, Germany.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3694885.3694886

iWOAR 2024, September 26–27, 2024, Potsdam, Germany

ACM ISBN 979-8-4007-1225-8/22/09
https://doi.org/10.1145/3694885.3694886

1 INTRODUCTION

Human Activity Recognition (HAR) is a rapidly growing field of research with applications in areas such as healthcare, sports, and security. The present work settles within the context of digital phenotyping of Autism Syndrome Disorder (ASD) [1] [2] [3] [4] which exploits clinical research, IoT, and computer vision to analyze children's play behavior and the motor peculiarities related to autism syndrome from different points of view: manipulation of the toy, posture, body movement, and exploratory activity of the surroundings. Many pedagogists and psychologists recognize that play is a fundamental way of learning [5]. Many models have been proposed to describe the relationship between children's play and their development [6] [7] and it has already been assessed that neuro-developmental disorders, such as ASD, affect the typical development of play behavior in children. Additionally, existing studies on children's behavior have a strong focus on children's social, emotional, and cognitive development [8], while little attention has been dedicated to the sensory-motor aspect of play [9]. In this context, our mission is to advance the SoA by simulating and studying children's play behaviors from the perspective of sensory-motor developments exploiting a non-invasive approach based on toys inertial measurements to support the early diagnosis of ASD, and to allow the anticipation of the intervention, which has been demonstrated to be more effective. From this standpoint, our current aim is to break down play behavior into sequences of fine-grained motor patterns with the intention of identifying and classifying some specific play actions or gestures. In our previous research, we investigated which play actions resulted as significant and measurable for the identification of ASD in target age of infants between 9 to 15 months [2]. The purpose of this paper is not to propose a specific methodology for ASD automatic play activity recognition but in turn we intend to demonstrate the feasibility of automatic recognition from a subset of the aforementioned actions exploiting an innovative fine-grained HAR approach.

Traditional methods for HAR often rely on sensor data from wearable devices or cameras, which can be used singularly or in parallel exploiting sensor fusion techniques, although using multiple sensors adds complexity in synchronizing different sources of data. In recent years, HAR deep learning approaches have grown in interest, despite the large amounts of labeled data and high computational power required, since they outperformed old machine learning-based approaches [10] [11] [12]. The present work's operational pipeline (Section 2) relies on inertial data to estimate positions that an object occupies in space exploiting the dead-reckoning approach. Dead reckoning is a fundamental process in navigation where external references, such as landmarks or GPS signals are unavailable. For example, it is crucial for Unmanned Aerial Vehicles (UAVs) because it can be used as a control technique assuming the current position estimate as accurate and calculating new control signals based on this estimate relative to a desired position in situations where GPS signal is lost [13]. For HAR purposes, this approach has not been deeply investigated in literature due to its challenging and intrinsic drift over time. Drift error can occur due to various factors such as changes in the user's walking speed, turning angles, or environmental conditions. These factors can cause the inertial sensors to accumulate errors over time, leading to inaccuracies in the estimated position and ultimately affecting the accuracy of the activity recognition. To mitigate the effects of drift error, various methods have been proposed in literature. These include sensor fusion techniques that combine multiple sensor modalities such as inertial measurement systems (IMUs) and GPS or cameras, when available, to improve the accuracy of the estimation. Alternatively, Kalman filters have been proposed to model the drift error and correct it in the estimation process. An example of usage of Kalman filters for HAR purposes can be found in [14], where the authors introduced a novel algorithm based on a quaternion representation, allowing accelerometer and magnetometer data to be used in an analytically derived and optimized gradient descent algorithm to compute the direction of the gyroscope measurement error as a quaternion derivative. We embraced the latter approach focusing our study on a single IMU sensor, whose signals are filtered and processed in order to obtain relative positions to feed neural networks and perform the classification task.

Section II describes in detail the workflow followed in the present study, with a comprehensive explanation of all its steps. Section III contains the results obtained and contextualizes them accounting also for the time-dependent nature of the signals. Finally, Section IV summarizes the findings of our work, expresses the limitations encountered, and provides some insight for future studies.

1.1 State of the Art

Different approaches can be found in the literature and many of them explored HAR tasks through deep learning models. Some methods rely on image processing. For instance, Ito et al. in [15] use acceleration and gyroscope values to compute a Fast Fourier Transform obtaining images used to feed a Convolutional Neural Network to perform classification. Instead, [16] aims to provide health monitor to patients through HAR. They started from raw acceleration data and exploited continuous wavelet transform to compute 2D images used as input of a Convolutional Neural Network.

There are also deep learning approaches based on raw signals such as [17], which fed a Long Short Term Memory network with raw accelerometer data to perform classification on WISDM Lab public dataset which includes coarse-grained activities (walking, standing, and jogging). Further, [18] forewent recurrent architectures and proposed a self-attention-based network on raw accelerometer data evaluating the network on different public datasets.

Other approaches are based on object tracking aiming at predicting the 3D position of an object with time. Several works involve pedestrian tracking through IMU sensors. This scenario introduces the possibility of using algorithms exploiting the intrinsic periodical frequency of the movements. For example, Fourati et al. in [19] introduced a foot-mounted Zero Velocity Update-aided IMU filtering algorithm for indoor pedestrian tracking. This algorithm assumes that the traced object (foot) assumes periodically a static position that can be used to adjust drifting errors. Hou et al. in [20] suggested that the head is a valid alternative to place the sensor on for pedestrian tracking, in fact, they provided a specific Pedestrian Dead-Reckoning method designed for head-mounted sensors. This method belongs to the field of Step-and-Heading Systems and aims to detect each step of the pedestrian by estimating its length and direction. Finally, it integrates every step to obtain a complete trajectory.

When the problem does not involve periodical movement, sensor fusion techniques are used, combining IMU and other long-term reliable sensors. Toy et al. in [21] used an improved Dead-Reckoning localization system using IMU sensor to improve Global Navigation Satellite System-based vehicle localization when the satellite signal is denied.

Brossard et al. in [22] proposed a method to track vehicles based only on IMU sensors. They used a Kalman Filter and a neural network to dynamically adapt the noise parameter. They evaluate the method on the KITTI odometry dataset reaching performances comparable to top-ranked methods which, by contrast, use LiDAR or stereo vision. Very few works about fine-grained HAR performed through not-worn sensors can be found in literature [3] [23].

The present work is focused on the development and implementation of a dead-reckoning-based system for human activity recognition. Dead-reckoning, a technique commonly used in navigation and robotics, will be employed to estimate the orientation and position of IMU-embedded toys (a car and a ball). The inertial data will then be processed and analyzed to recognize and classify various fine-grained human activities related to the play.

The goals of this work are: i) To demonstrate a methodology which enhances the accuracy and efficiency of HAR systems, particularly in environments where traditional methods, such as GPS, may be limited or unavailable; ii) To prove the feasibility of an approach which could be applied to ASD children, that are characterized by hypersensitivity and thus intolerance to wearable sensors. Moreover, the proposed approach does not rely on any computer vision technique, thus eliminating the necessity of cameras that record the space the subject is moving into. The outcomes of this research will potentially have a significant impact on fields ranging from healthcare and fitness monitoring to augmented and virtual reality applications.

2 METHODS

The workflow followed for our purpose has been divided into three major tasks:

(1) Dead-reckoning, to estimate position from inertial raw data
(2) Feature extraction
(3) Classification through neural networks

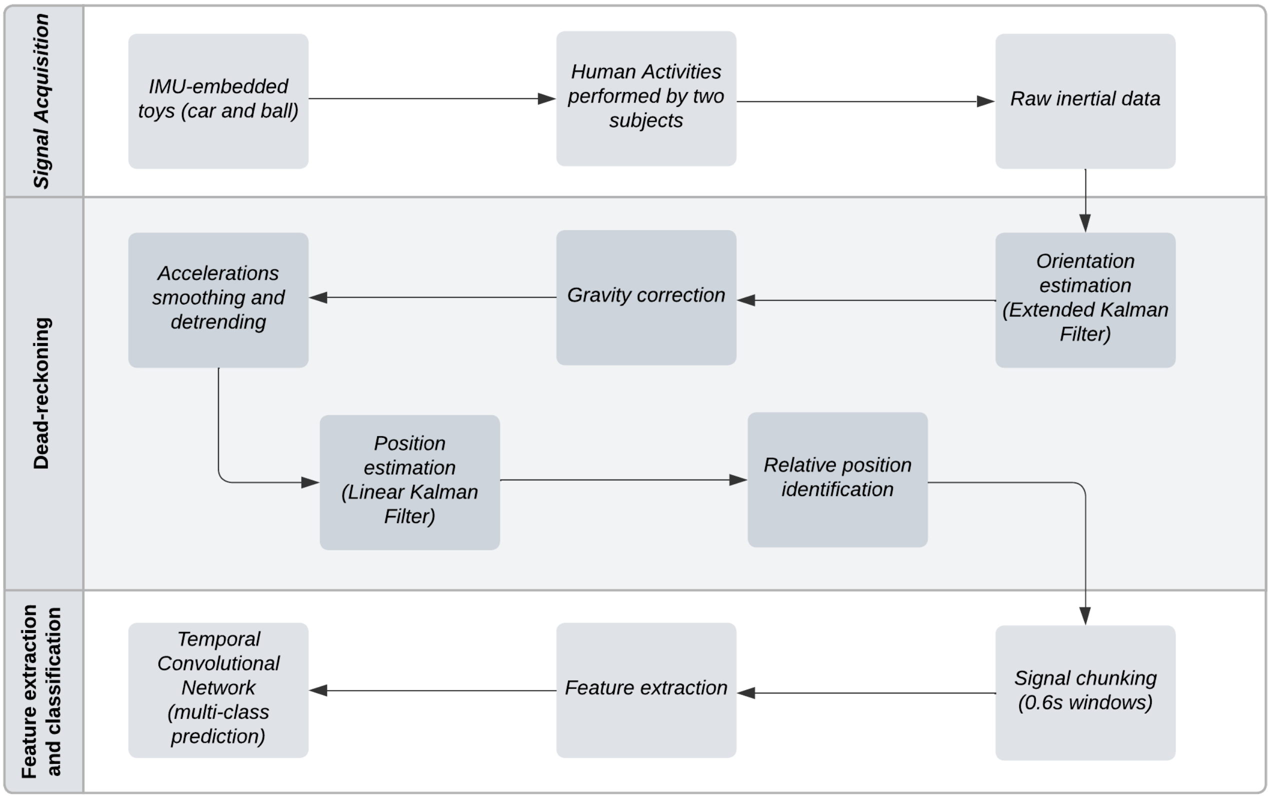


Figure 1: Detailed workflow.

2.1 Dead-Reckoning

In navigation, dead-reckoning is the process of estimating the position of a moving object from IMU sensor only.

Dead-reckoning is a kind of path integration: by integrating the acceleration of an object twice, its position can be obtained. IMU sensors provide angular velocity and acceleration. The latter term incorporates a constant gravity component we want to get rid of. Since we do not know gravity's components projections over the three axes of the IMU but only its magnitude, to retrieve the acceleration of the object it is necessary to find the orientation of the sensor and rotate it in the reference plane to isolate and remove gravity from the vertical axis, then rotate the sensor again and integrate the acceleration twice (see Figure 2).

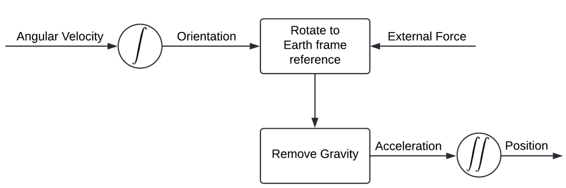


Figure 2: Dead-reckoning pipeline: inertial measurements are integrated to obtain position and orientation [24].

Theoretically, dead-reckoning seems to work perfectly but, practically, the computation of integrals hides some challenges. In fact, the process involves three integrals: one integral for orientation estimation and two subsequent integrals for position estimation. Since the input of the integrals are sensor measurement values, the measurement error is carried into the computation and it grows in a nonlinear way. Moreover, the double integration for position estimation carries the integration constant, thus leading to signal drift. For this reason, tracking becomes infeasible (see Figure 3).

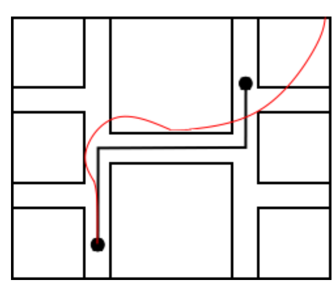


Figure 3: Unbounded growing error on simple path tracking through dead-reckoning double integration [25].

The explored approach replaces orientation estimation with an Extended Kalman Filter and position estimation with a Linear Kalman filter. Kalman Filters are Hidden Markov Models that produce estimates of hidden and observed variables by predicting the next values merging an estimate of the following value with its effective measured value. The estimate takes into account the noise of the process and the noise in the measurement values. The dead-reckoning process takes advantage of Kalman filters by reducing its intrinsic drift error. The detailed pipeline is shown in Figure 1. From IMU-embedded toys, raw acceleration values are collected and go through an Extended Kalman Filter that returns an estimation of the orientation. Orientation estimation helps obtain the object acceleration since the gravity component can be easily removed once the sensor has been rotated into the earth frame reference, where the gravity component is known. Then acceleration is smoothed through a moving average filter and detrended before going into a Linear Kalman Filter to obtain position estimation which can be turned into relative positions.

This part of the pipeline manages the limitation of the drift error, by applying a form of correction: every position is transformed into the relative position vector related to the previous position (see Figure 4).

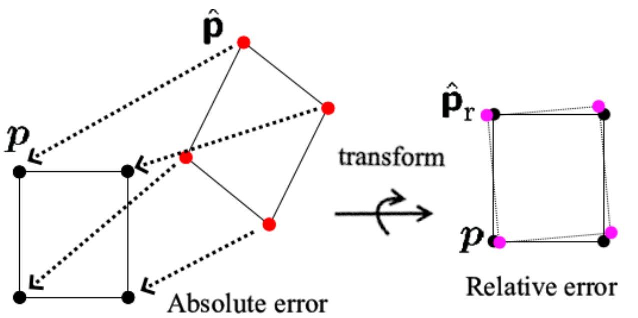


Figure 4: Comparison of absolute and relative errors [26].

Processing is possible because the aim of the project is not tracking but activity recognition. By transforming each position into a relative position vector related to the previous position, the pipeline mitigates the effect of drift error. This approach is sufficient for activity recognition since the relative changes in position can still reflect the characteristics of different activities. Absolute position accuracy is not crucial in this context. What matters is the relative movement patterns, which can still be accurately derived despite some level of drift or position error.

To correctly initialize the process noise covariance matrix of the Extended Kalman Filter we used the noise values along the 3-axes provided by the Shimmer3 IMU datasheet.

2.2 Features extraction

The following step is to generate an *instantaneous displacement signal* where each sample represents the differential position with respect to the previous one. Finally, the signal is segmented into 0.6 seconds-long non-overlapping windows used as temporal support to build a signal representation suitable for neural networks. In particular, each window was transformed into an image in order to feed a particular type of convolutional neural network as explained in Section 2.3. Each image is built by stacking the three axial-vector components one upon the other (see Figure 5).

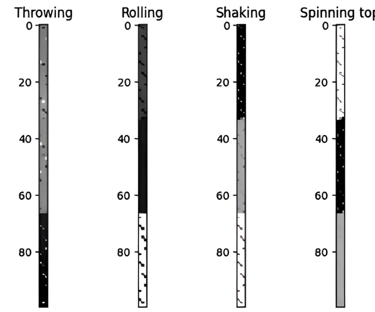


Figure 5: Chunked signal image representation.

2.3 Model selection and training

In [23], it was demonstrated that traditional machine learning approaches perform well on coarse-grained activity recognition but showed also a consistent performance drop when applied to fine-grained activity recognition. In this study, instead, we exploited a deep learning approach based on a Temporal Convolutional Network (TCN), a particular type of Convolutional Neural Network (CNN) that was first proposed by Lea et al. in [27] capable of capturing both spatial-temporal features like a CNN and high-level temporal information as a Recurrent Neural Network (RNN). Figure 6 presents the architecture of the model. A single TCN layer (number of filters = 108; kernel size = 6; dilations = 1,2,4 and 8; activation

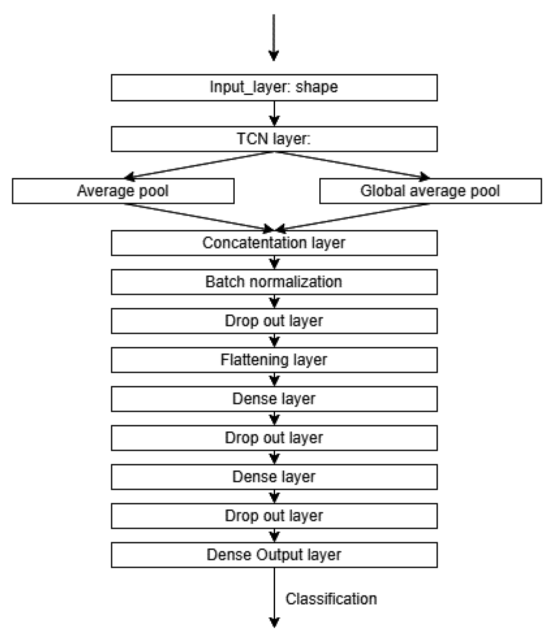


Figure 6: TCN model architecture capable of capturing spatial-temporal features.

function = ReLU) output flows into two parallel pooling functions and then is concatenated. Batch normalization is applied and the model is regularized through dropout. Then, there is a flattening layer and a series of dense layers until the output passes through a softmax function that returns a vector of probabilities.

2.3.1 The dataset. A synthetic dataset has been collected using Shimmer3© IMU sensors by two different subjects (adults) simulating the play behavior of children with two different toys: a ball and a car. For each toy, a set of activities has been investigated (see Table 1), particularly the ones that resulted in the most frequently chosen by children in a previous study [2]. Each activity was recorded for 30 seconds from both subjects at a sample rate of 100 Hz. While the sensor embedded in the ball was placed inside of it, instead, the sensor embedded in the car was placed inside one of its wheels, in order to be able to differentiate among different fine-grained activities. These activities have been performed in an indoor environment.

Table 1: Performed activities divided by toy.

Ball			
Throw	Roll	Shake	Spinning top
Car			
Flip	Slide	Hit	Spin wheels

2.3.2 Training methodology. A training set (TRS) and a test set (TS) were derived from the original dataset, which was composed of 2’396 observations. In particular, TRS contained 75% of observations and TS contained the remaining 25%. In order to account for the temporal nature of the data, it was adopted a custom splitting approach which did not contemplate any shuffle of the data. It was decided to divide each window into two portions: the first one assigned to TRS and the second one assigned to TS. The network has been separately trained on the two toys using cross-entropy and ADAM as loss function and optimizer, respectively. The training has been done on 300 epochs using a batch size of 32. All the exploited hyperparameters have been properly tuned.

The code associated with this project is accessible via Zenodo at the following link: https://zenodo.org/doi/10.5281/zenodo.13623602.

3 RESULTS AND DISCUSSION

3.1 Results on car toy

The car toy-related fine-grained activities are four: spinning wheels, knocking, flipping, and sliding (typical activities performed by Typical Development - TD and Not Typical Development - NTD children). Results are shown in Figure 8. The network reaches a training accuracy of 89% and a test accuracy of 69%. From the confusion matrix, it can be noticed that the network is stronger in predicting ’hit’ activity but it does not meet performance expectations when it has to differentiate between ’slide’ and ’spin wheels’.

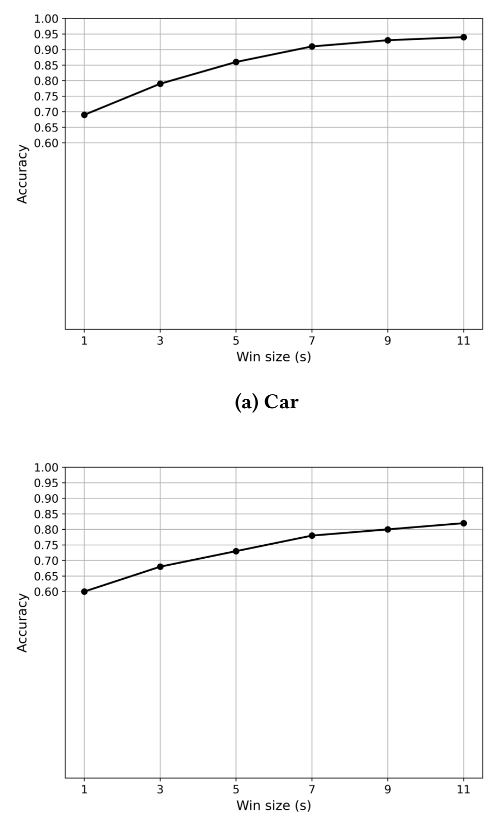
3.2 Results on ball toy

The ball toy-related fine-grained activities are four: throwing, rolling, shaking, and spinning top. Results are shown in Figure 9. The network reaches a training accuracy of 83% and a test accuracy of 60%. From the confusion matrix, it can be noticed that the network is more accurate in predicting ’shake’ but it struggles in differentiating between ’roll’ and ’spin’ (spinning top).

As we expected, recognizing ball activities is harder than recognizing car ones because the sensor is placed inside the ball and the activities of rolling the ball and spinning the ball from the top are both based on a rotation of the toy but, while during the first the ball moves through space, instead, during the second it rotates with a fixed contact point on a surface.

3.3 Exploiting time-series continuity for predictions

The test accuracy is the canonical metric used for performance evaluation. It is the sum of all correct predictions over the total samples. Considering activities as chunks of signals completely independent from each other can be misleading because we lose the concept of temporal continuity of time series. We can exploit



(a) Car

(b) Ball

Figure 7: Continuity validation method. Test accuracy increases as the window size increases.

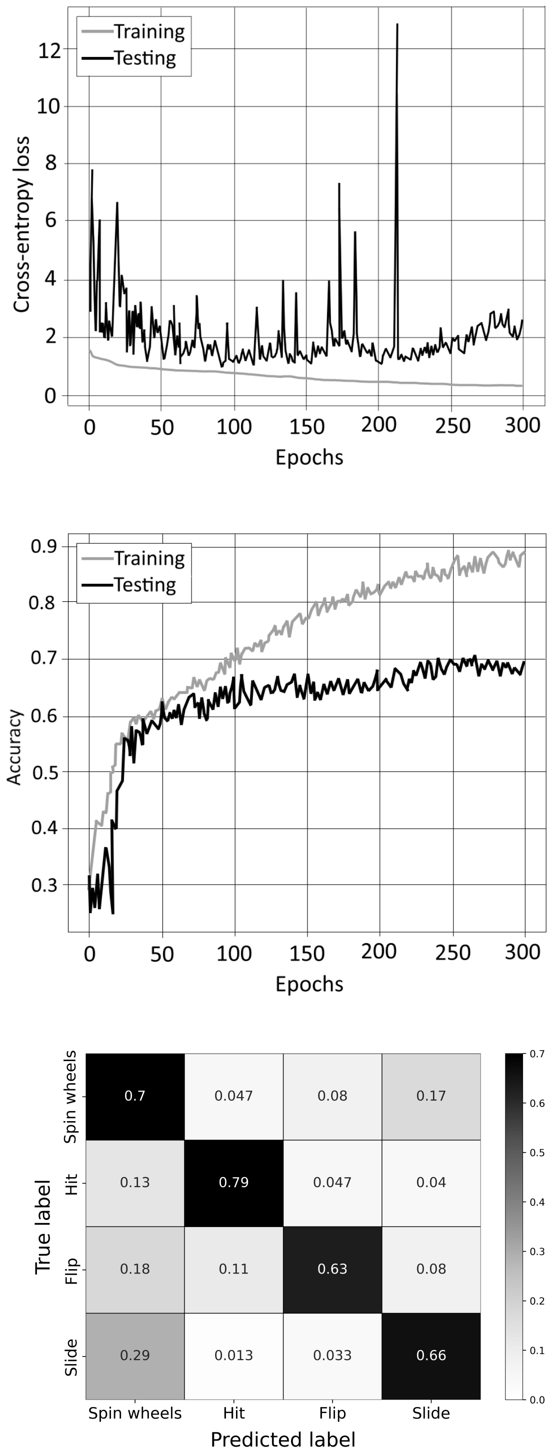


**Figure 8: Results on car toy.
TOP Cross-entropy loss.
MIDDLE Model accuracy.
BOTTOM Confusion matrix.**

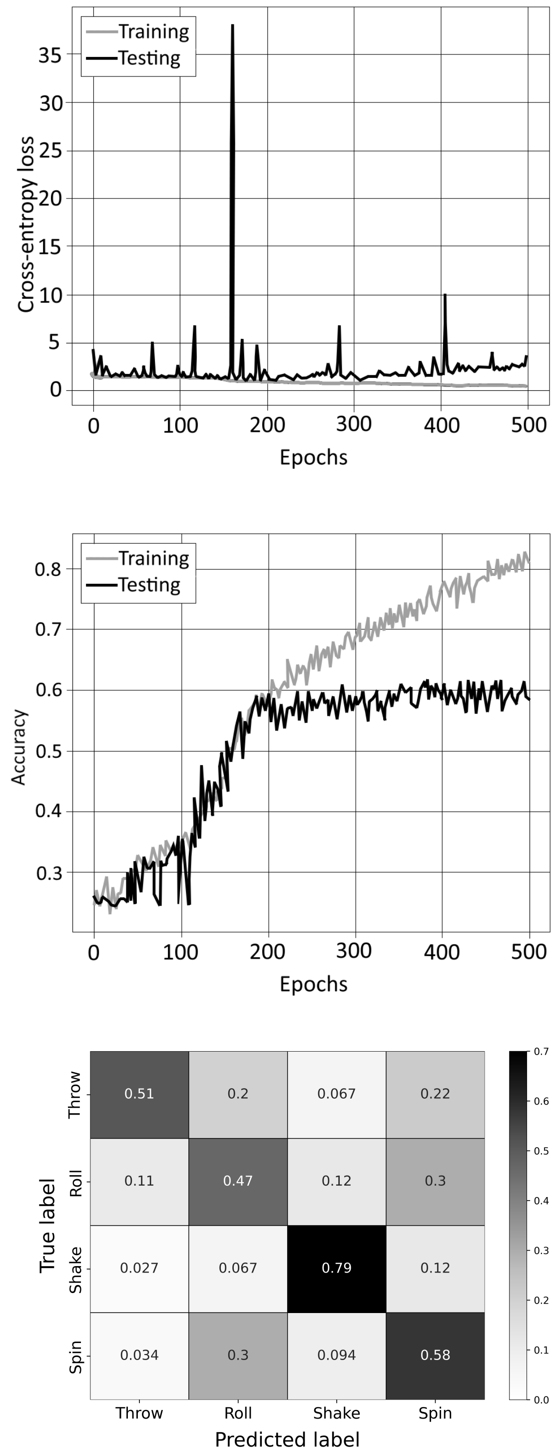


**Figure 9: Results on ball toy.
TOP Cross-entropy loss.
MIDDLE Model accuracy.
BOTTOM Confusion matrix.**

the intrinsic continuous nature of the signal for the prediction task. The idea is that for the i^{th} chunk, the prediction is the most present value in a window centered over it. For instance, by taking advantage of this simple but effective technique it is possible to correct a prediction error caused by an occurrence of some activity, namely 'B', predicted within a sequence of another activity, namely 'A', in a certain period of time.

Figures 7a and 7b clarify how this simple practice significantly improves the performance of the proposed approach. On the y-axis there is the accuracy value while on the x-axis there is the window size. A window size of 1 stands for looking only at the raw predictions of the model, while the next values use the validation method explained before.
There is a positive trend in accuracy that tends to increase with the window size, in fact, the model has less possibility to return an error when it also considers the predictions in the chunk before and after the current one. This is true because the model is not random guessing. Given a certain increment in window size, the accuracy improvement resulted higher if the baseline accuracy was higher. In fact, given a window size of 3, the network trained for the car increased about 10% in test accuracy against the baseline, while the one trained on the ball increased about 8%. A window size of 3 is already promising, and larger sizes appear to be even more performant. However, longer window sizes result in higher performance only when the child engages in a specific play activity for a longer time. Such a case becomes increasingly implausible as the size of the window increases.

Summing up, it is clear how considering a wider window for activity predictions helps the proposed approach to identify the true activity in the current chunk.

4 CONCLUSIONS

The present work wants to be an alternative approach for fine-grained human activity recognition based on dead-reckoning and temporal convolutional networks. The proposed method aims to extract features from inertial data applying Kalman filter-corrected dead-reckoning and accurately classifying specific actions or gestures performed by individuals. The novelty of the present approach relies on the combination of Kalman filters and neural networks to successfully accomplish the HAR task using the former to mitigate errors and transform absolute positions into relative ones to train the latter. Moreover, for the present work the inertial dataset has been collected through smart toys (IMU-embedded toys) thus differing from many wearable-based approaches present in literature [28] [29] [30], which seem to outperform our approach reaching even more than 90% accuracy. However, this is an unfair comparison because of two main reasons: (i) many works in literature are based on non-hand-oriented activities, such as walking or going upstairs, which are intrinsically simpler to recognize with respect to the hand-oriented activities this works focuses on; (ii) the nature of data itself. In fact, wearable devices worn by a subject produce more deterministic inertial data than IMU-embedded toys since they are constrained to move with the subject himself. In layman's terms, these devices will produce comparable signals if the same activity is performed, while this assumption is not always true for IMU-embedded toys. In conclusion, our approach shows the potential of combining dead-reckoning and temporal convolutional networks for fine-grained human activity recognition in a different domain with respect to other related works. The presented approach can find many applications related to the broad spectrum of ASD. For instance, the identification of "lower-order" motor repetitions which include toys manipulation, repetitive play patterns, banging toys together, toy tranfer from one hand to the other, etc. [31][32]. From a clinical perspective, the last two belong to the milestones in children neurological development [33].

The main limitations of the presented approach rely on the reduced number of tasks analyzed, either in terms of activities and exploited toys. However, we remain confident that the present work can represent a solid base for future works in the perspective of developing IoT systems capable of supporting clinicians in the early diagnosis of ASD, and capable of distinguishing different phenotypes of autism with the intention of delivering patient-specific treatments [4]. Future works will focus on overcoming the aforementioned limitations. For instance, they could focus on further refining the dead-reckoning estimation process to reduce drift error and improve the accuracy of HAR or on training new networks for other activities in order to investigate more deeply the potentialities of this approach. Moreover, such a system could be integrated with computer vision techniques, such as the one presented in [28], in order to build multimodal diagnostic systems that provide new informative viewpoints and improve the treatment quality. The presented feasibility study assessed our methodology over data from adults, so, to further generalize it, we are currently collecting data to test the proposed pipeline over a wider and more variegated pool of subjects comprehending children with ASD.

ACKNOWLEDGMENTS

This research is supported by the AI4Autism project (Digital Phenotyping of Autism Spectrum Disorders in children, grant agreement no. CR- SII5 202235 / 1) of the Sinergia interdisciplinary program of the SNSF.

REFERENCES

[1] Faraci, F. D., Papandrea, M., Puiatti, A., Agustoni, S., Giulivi, S., D'Apuzzo, V., ... & Rossini, E. (2018, June). "AutoPlay: a smart toys-kit for an objective analysis of children ludic behavior and development". In 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE

[2] Rossini, E., Faraci, F., Papandrea, M., Puiatti, A., Bernaschina, S., & Ramelli, G. P. (2019, June). "AutoPlay: smart games for evolutionary screening under 14 months". In SWISS MEDICAL WEEKLY (Vol. 149, pp. 6S-6S). FARNSBURGERSTR 8, CH-4132 MUTTENZ, SWITZERLAND: EMH SWISS MEDICAL PUBLISHERS LTD

[3] Bonomi, Niko, & Michela Papandrea. "Non-intrusive and Privacy Preserving Activity Recognition System for Infants Exploiting Smart Toys." In EAI International Conference on IoT Technologies for HealthCare, pp. 3-18. Cham: Springer International Publishing, 2021.

[4] Tafasca, S., Gupta, A., Kojovic, N., Gelsomini, M., Maillart, T., Papandrea, M., ... & Odobez, J. M. (2023, October). The AI4Autism Project: A Multimodal and Interdisciplinary Approach to Autism Diagnosis and Stratification. In Companion Publication of the 25th International Conference on Multimodal Interaction (pp. 414-425).

[5] Rodger, S., & Ziviani, J. (1999). Play-based occupational therapy. International Journal of Disability, Development and Education, 46(3), 337-365.

[6] Page, J., Nutbrown, C., & Clare, A. (2013). Working with babies and children: From birth to three. Sage.

[7] Mulligan, S., & White, B. P. (2012). Sensory and motor behaviors of infant siblings of children with and without autism. The American Journal of Occupational Therapy, 66(5), 556-566.

[8] Carpenter, K. L., Hahemi, J., Campbell, K., Lippmann, S. J., Baker, J. P., Egger, H. L., ... & Dawson, G. (2021). Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism. Autism Research, 14(3), 488-499.
[9] Suzuki, S., Amemiya, Y., & Sato, M. (2019, October). Enhancement of gross-motor action recognition for children by CNN with OpenPose. In IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society (Vol. 1, pp. 5382-5387). IEEE.
[10] Sharma, V., Gupta, M., Pandey, A. K., Mishra, D., & Kumar, A. (2022). "A review of deep learning-based human activity recognition on benchmark video datasets". Applied Artificial Intelligence, 36(1), 2093705.
[11] Gupta, S. (2021). "Deep learning based human activity recognition (HAR) using wearable sensor data". International Journal of Information Management Data Insights, 1(2), 100046.
[12] Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). "A review of human activity recognition methods". Frontiers in Robotics and AI, 2, 28
[13] Power, W., Pavlovski, M., Saranovic, D., Stojkovic, I., & Obradovic, Z. (2020). Autonomous navigation for drone swarms in GPS-denied environments using structured learning. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 219-231). Springer International Publishing.
[14] Madgwick, S. O., Harrison, A. J., & Vaidyanathan, R. (2011, June). Estimation of IMU and MARG orientation using a gradient descent algorithm. In 2011 IEEE international conference on rehabilitation robotics (pp. 1-7). IEEE.
[15] Ito, C., Cao, X., Shuzo, M., & Maeda, E. (2018, October). "Application of CNN for human activity recognition with FFT spectrogram of acceleration and gyro sensors". In Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers (pp. 1503-1510).
[16] Nedorubova, A., Kadyrova, A., & Khlyupin, A. (2021). "Human activity recognition using continuous wavelet transform and convolutional neural networks". arXiv preprint arXiv:2106.12666.
[17] Chen, Y., Zhong, K., Zhang, J., Sun, Q., & Zhao, X. (2016, January)." LSTM networks for mobile human activity recognition". In 2016 International conference on artificial intelligence: technologies and applications (pp. 50-53). Atlantis Press.
[18] Mahmud, S., Tonmoy, M., Bhaumik, K. K., Rahman, A. M., Amin, M. A., Shoyaib, M., ... & Ali, A. A. (2020). "Human activity recognition from wearable sensor data using self-attention". arXiv preprint arXiv:2003.09018.
[19] Fourati, H., Manamanni, N., Afilal, L., & Handrich, Y. (2013, July). "Position estimation approach by complementary filter-aided IMU for indoor environment". In 2013 European Control Conference (ECC) (pp. 4208-4213). IEEE.
[20] Hou, X., & Bergmann, J. (2020). "A pedestrian dead reckoning method for head-mounted sensors". Sensors, 20(21), 6349.
[21] Toy, I., Durdu, A., & Yusefi, A. (2022, November). "Improved Dead Reckoning Localization using IMU Sensor". In 2022 International Symposium on Electronics and Telecommunications (ISETC) (pp. 1-5). IEEE.
[22] Brossard, M., Barrau, A., & Bonnabel, S. (2020). "AI-IMU dead-reckoning". IEEE Transactions on Intelligent Vehicles, 5(4), 585-595.
[23] Pandurangan, Shalini, Michela Papandrea, & Mirko Gelsomini. "Fine-Grained Human Activity Recognition - A new paradigm." Proceedings of the 7th International Workshop on Sensor-based Activity Recognition and Artificial Intelligence. 2022.
[24] Huang, C. J., Chi, C. J., & Hung, W. T. (2023). "Hybrid-AI-Based iBeacon Indoor Positioning Cybersecurity: Attacks and Defenses". Sensors, 23(4), 2159.
[25] Fallah, N., Apostolopoulos, I., Bekris, K., & Folmer, E. (2012, May). "The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 425-432).
[26] Liu, Y., Wang, Y., Wang, J., & Shen, Y. (2020). "Distributed 3D relative localization of UAVs". IEEE Transactions on Vehicular Technology, 69(10), 11756-11770.
[27] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). "Temporal convolutional networks for action segmentation and detection". In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 156-165).
[28] Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H. N., Butt, A., & Bashir, A. K. (2022). A review of machine learning-based human activity recognition for diverse applications. Neural Computing and Applications, 34(21), 18289-18324.
[29] Hou, C. (2020, May). A study on IMU-based human activity recognition using deep learning and traditional machine learning. In 2020 5th International Conference on Computer and Communication Systems (ICCCS) (pp. 225-234). IEEE.
[30] Zhuang, W., Chen, Y., Su, J., Wang, B., & Gao, C. (2019). Design of human activity recognition algorithms based on a single wearable IMU sensor. International Journal of Sensor Networks, 30(3), 193-206.
[31] Caldwell-Harris, C. L. (2021). An explanation for repetitive motor behaviors in autism: facilitating inventions via trial-and-error discovery. Frontiers in Psychiatry, 12, 657774.
[32] Tian, J., Gao, X., & Yang, L. (2022). Repetitive restricted behaviors in autism spectrum disorder: from mechanism to development of therapeutics. Frontiers in neuroscience, 16, 780407.
[33] Dosman, C. F., Andrews, D., & Goulden, K. J. (2012). Evidence-based milestone ages as a framework for developmental surveillance. Paediatrics & child health, 17(10), 561-568.

Received 13 June 2024; Revised 8 August 2024; Accepted 4 September 2024