

CLADSI: Deep Continual Learning for Alzheimer’s Disease Stage Identification Using Accelerometer Data

Santos Bringas ^{1b}, Rafael Duque ^{1b}, Carmen Lage ^{1b}, and José Luis Montaña ^{1b}

Abstract—Alzheimer’s disease (AD) is a neurodegenerative disorder that can cause a significant impairment in physical and cognitive functions. Gait disturbances are also reported as a symptom of AD. Previous works have used Convolutional Neural Networks (CNNs) to analyze data provided by motion sensors that monitor Alzheimer’s patients. However, these works have not explored continual learning algorithms that allow the CNN to configure itself as it receives new data from these sensors. This work proposes a method aimed at enabling CNNs to learn from a continuous stream of data from motion sensors without having full access to previous data. The CNN identifies the stage of AD from the analysis of data provided by motion sensors. The work includes an experimentation with data captured by accelerometers that monitored the activity of 35 Alzheimer’s patients for a week in a daycare center. The CNN achieves an accuracy of 86,94%, 86,48% and 84,37% for 2, 3 and 4 experiences respectively. The proposal provides advantages to working with a continuous stream of data so that the CNN are constantly self-configuring without the intervention of a human. The work can be considered as promising and helpful in finding deep learning solutions in medical cases in which patients are constantly monitored.

Index Terms—Alzheimer’s disease, continual learning, convolutional neural network.

I. INTRODUCTION

ALZHEIMER’S disease (AD) is a neurodegenerative disorder characterized by loss of memory and progressive cognitive impairment. This disease is a public health problem affecting 55 million people worldwide [1]. AD can lead to a significant impairment in these people’s physical functions and a loss of independence in basic activities of daily living (use of the bathroom, dressing, etc.) [2]. Gait disturbances of different characteristics are also reported as a symptom of AD [3]. Thus, Varma et al. [4] point out that variability in step velocity and cadence can predict individuals with early and mild AD.

Wearable motion sensors (accelerometers, gyroscopes, etc.) have been used to measure gait variables when AD patients perform controlled activities such as walking 40 meters [5]. These measurements can be carried out in human motion laboratories using multiple sensors that provide complete and accurate gait information [6].

Smartphone sensors can permanently capture motion data in the places where the patient carries out their daily activities without using motion analysis laboratories. The amount and frequency of collected data by these smartphone sensors led us to explore the feasibility of a method that learns from this stream of data and identifies the stage of AD. Thus, a continuous data stream refers to a constant flow or sequence of data points or information that occurs over time without interruption. A continual learning algorithm or model [7] aims to learn from this continuous stream of data, given in several parts, and without having access to previously seen data, or only to a small amount of said data. In the context of machine learning and artificial intelligence, a task denotes a specific objective or goal that a system aims to accomplish. It typically involves processing input data to produce desired outputs or predictions. Tasks can vary widely in complexity and nature, ranging from simple classification or regression problems to more intricate tasks such as natural language processing, image recognition, or reinforcement learning. Each task typically requires its own set of algorithms, techniques, and evaluation metrics tailored to its particular requirements and objectives. The specific task of this work involves identifying the state of AD using accelerometer data. Thus, this paper proposes applying the continual learning

Manuscript received 30 June 2023; revised 15 March 2024; accepted 14 April 2024. Date of publication 22 April 2024; date of current version 6 June 2024. The work of Santos Bringas was supported by the University of Cantabria, Government of Cantabria and Banco Santander through an industrial doctorate under Grant DI27, awarded in the 2020 Industrial Doctorate Program. The work of Rafael Duque and José Luis Montaña was supported in part by the “Proyectos Puente 2021” from the Consejería de Universidades, Igualdad, Cultura y Deporte, Government of Cantabria, under Grant 21.VP50.64662, and in part by the Consejería de Universidades, Igualdad, Cultura y Deporte, Government of Cantabria, under Grant SUBVTC-2023-0013. (Corresponding author: Rafael Duque.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee for Clinical Research of Cantabria and the Ethics Committee for Research Projects of the University of Cantabria.

Santos Bringas was with the Department of Artificial Intelligence, Axpe Consulting Cantabria, 39600 Camargo, Spain (e-mail: santos.bringas@unican.es).

Rafael Duque and José Luis Montaña are with the Faculty of Science, Department of Mathematics, Statistics and Computer Science, University of Cantabria, 39005 Santander, Spain (e-mail: rafael.duque@unican.es; joseluis.montana@unican.es).

Carmen Lage is with the Global Brain Health Institute, University of California, San Francisco, CA 94143 USA (e-mail: carmen.lage@gbhi.org).

Digital Object Identifier 10.1109/JBHI.2024.3392354

paradigm so that a deep learning model can identify the stage of AD from the analysis of a patient's gait. The validation of the study is conducted utilizing a dataset containing information from 35 patients with AD who were monitored for a week at a daycare center via mobile phone accelerometers. The results are then compared with those produced by a Convolutional Neural Network (CNN) trained using non-continual learning.

This article includes seven sections. Section II studies background in the field of continual learning and reviews works that have analyzed the gait of AD patients. Section III studies the feasibility of the continual learning paradigm to identify AD stage from patients' gait data. Section IV shows a case study that applies this new method. Section V discusses the results of applying this proposal. Section VI studies the threads to the validity of the results. Section VII analyses the conclusions of this work.

II. BACKGROUND

This section presents theoretical content and previous work on which this research proposal is based. First, the continual learning paradigm is analyzed. Secondly, the section includes an analysis of the evidence and works that relate AD and gait analysis.

A. Continual Learning

One significant challenge in artificial neural networks is catastrophic forgetting [8]. This issue refers to a phenomenon observed in connectionist networks, where new learning disrupts or erases previously acquired knowledge. This occurs particularly in networks trained sequentially, as new learning alters the connection weights involved in representing old learning. The term catastrophic emphasizes the significant and often detrimental impact this forgetting has on the network's overall performance, particularly in specific network configurations.

While neural network training is designed to be concurrent, with data learned through repetition until an optimal state is reached, human learning is sequential. In human learning, several pieces of data are presented iteratively, and once an optimal state is achieved, the next piece is learned, serving as a bias for subsequent iterations, thereby facilitating faster learning compared to learning them separately.

Therefore, if we aim to use neural networks to solve tasks that require human-like understanding, they should ideally learn in a sequential manner similar to humans. However, catastrophic forgetting arises, hindering our ability to address this problem. Goodfellow et al. [9] conducted empirical studies demonstrating that, in several experiments, every tested network forgot the first task while learning a new one. These authors also attempted to address this issue using older techniques such as Dropout or activation functions but did not achieve significant results, as they are neither specifically oriented nor optimal for addressing this matter.

This issue may seem evident, as neural networks' optimization algorithms attempt to fit the training data to minimize prediction error by adjusting weights. However, their primary objective is to learn relevant information from a fraction of the data (training partition) to generalize and predict the entire dataset (validation and test partitions) and future inputs. They

employ different processes to reduce overfitting on the training data. Nevertheless, if the training set changes, the information previously learned from the previous sets will be forgotten due to the inherent nature of the learning system described above.

This complexity in retaining memory as a neural network learns tasks iteratively highlights the limitations of neural networks in learning over time, necessitating a thorough solution to this problem. Continual learning algorithms present promising solutions, aiming to enable the network to retain learned information and mitigate these effects, allowing for continuous learning over time and adaptation to potential changes in the initial problem set. Generally, continual learning algorithms are broadly categorized into the following three classes or method groups [10]:

- *Replay methods* involve repeating representative past data stored in a small memory or generating synthetic data from learned features. When learning new tasks, stored examples or synthetic ones from previous tasks are fed to the network to prevent forgetting. Some of these methods utilize stored data to guide the optimization process of the new task by analyzing the data.
- *Regularization methods* attempt to adapt the network weights to balance new information and learned features. These algorithms adjust the loss function to balance learned information from previous tasks with information to be learned from the current one.
- *Parameter isolation methods* involve modifying the network architecture or restricting certain parts for specific tasks when necessary. Some methods expand the network to allocate a new part to the new task, while others freeze certain weights containing important information from one task to prevent forgetting or overwriting.

Some of the latest proposed methods have shown promising results in state-of-the-art datasets [11], [12]. Therefore, exploring these alternatives is advisable, as they allow for the introduction of data in small increments, facilitating gradual learning and progressive improvement of results, and even enabling the introduction of new classes or tasks to the network. These methods enable lifelong learning but there still challenges such as catastrophic forgetting and increased complexity to be faced. Moreover, the ethical debate concerning the integration of continual learning algorithms in medicine can revolve around their capacity to autonomously adapt with minimal human intervention. Striking a balance between harnessing the potential advantages of these algorithms and ensuring appropriate human oversight remains crucial.

B. Deep Learning and Alzheimer's Disease

Beauchet et al. [13] perform a meta-analysis of twelve studies that concludes that poor gait performance predicts non-AD and AD dementia. According to Ardlie et al. [14], there is early evidence of discrete pathological signatures of gait in very mild AD because of the disease-specific role of cognition in gait. Even recent studies suggest that measured gait performances can be potential digital biomarkers of cognitive frailty [15]. This has motivated the use of methods based on artificial intelligence to perform gait analysis and identify cognitive impairment. Alharthi et al. [16] review research works that apply artificial

TABLE I
ANALYSIS OF DATA FROM MOTION SENSORS IN PATIENTS WITH ALZHEIMER'S

Study	Technique	Device	Tasks of patients	Goal	Results
Ghoraani et al. [18]	Support vector machine	Computerized walkway	Dual-task gait	Detection of mild cognitive impairment and AD	88% accuracy (90% F1-score)
Mc Ardle et al. [14]	Statistical analysis	A single accelerometer based wearable	Six 10-meter walks	Differentiate dementia disease subtypes	Statistical evidence that wearable technology has potential as a clinical tool for differential diagnosis
You et al. [19]	Long Short-Term Memory (LSTM)	Kinect V2	Walk the length of a 10 meter-path and back	Diagnosis of mild cognitive impairment and AD	90.48% accuracy
Gillain et al. [20]	Mann-Whitney U test	An accelerometer in the patient's belt	Walks along a straight 40-meter corridor and back again	Diagnosis of mild cognitive impairment in people who develop AD	Gait speed, symmetry and regularity were lower in MCI than developed AD
Pedrero-Sánchez et al. [21]	Multi-head convolutional layered neural network	Gyroscope and accelerometer attached to the participant's back	Five specific tasks with arms and legs	Classification of healthy, Alzheimer and Parkinson population	100% of the participants are correctly assigned to each group

intelligence techniques for human gait analysis and conclude that deep learning CNNs usually outperform shallow learning models. Seifollahi et al. [17] carried out a study in which 70 elderly subjects with mild cognitive impairment participate along with 80 elderly control subjects with good health. This study collects data from a Kinect device and uses an adaptive Neuro-Fuzzy Inference System to identify mild cognitive impairment with values of more than 90% for accuracy, sensitivity, and F-score.

Statistical analysis is employed by some researchers [14], [20] without resorting to machine learning for model generation. Conversely, other studies [18], [19], [21] utilize machine learning techniques, although they do not adopt the continual learning paradigm. The objectives of these studies are summarized in Table I. For instance, Mc Ardle et al. [14] carry out a study into how gait analysis can differentiate dementia disease subtypes. Additionally, accelerometer data is used by another set of studies [18], [19], [20] to diagnose mild cognitive impairment related to AD. Lastly, Pedrero-Sánchez et al. [21] develop a CNN to process gyroscope and accelerometer data, classifying populations into healthy, Alzheimer's, and Parkinson's groups. However, these studies do not address the identification of the stage of AD, which is the focus of this work.

It should be noted that usually all the works monitor tasks in which the patient walks a few meters and do not monitor daily activities over several hours. In addition, this monitoring is usually carried out by body computers attached to the patient's body [14], [20], [21] or devices that are not usually as accessible as a mobile phone [18], [19].

To conclude, it can be affirmed that there is clear evidence of the relationship between gait problems and AD. Numerous works have applied deep learning models to study this relationship by capturing data from sensors. However, a line of work is emerging that has not yet been explored in depth, and whose objective is that the deep learning models self-configure without the intervention of an expert in artificial intelligence as more data is collected from the sensors.

III. METHOD

This section presents a method of exploring the feasibility of applying the continual learning paradigm to identify the stage

of AD using data captured by the patient's accelerometer. The CLADSI system was developed to execute a continual learning process for stage identification of AD. Firstly, this section outlines the continual learning process carried out by CLADSI. Secondly, it illustrates the process initiation with a CNN, which will be introduced later. Finally, the section describes the procedure for implementing the remaining steps of the process, including the retraining of the CNN and the evaluation of results.

A. CLADSI

A-GEM (Attribution-based Gradient Episodic Memory) [22] is a continual learning algorithm designed to address the problem of catastrophic forgetting in neural networks. It belongs to the group of replay methods (see Section II-A) and aims to retain the information learned from previous tasks while adapting to new ones. A-GEM achieves this by storing a limited number of examples from previously seen data and constraining the optimization for a new task based on these stored examples. During training, a random sample of the stored examples is selected, and the optimization process ensures that the current task's data does not increase loss in these selected examples. This approach directs the neural network towards solutions that optimize both the current task and the previously learned ones. However, a potential drawback of A-GEM is that as more tasks are observed, it becomes increasingly challenging to learn new ones, potentially leading to optimization difficulties if there is no gradient vector to reduce loss in all selected samples.

CLADSI performs a process based on A-GEM algorithm since it allows for the input of a continuous accelerometer data stream of Alzheimer's patients. This method follows the following phases:

- 1) Initialization: The parameters of the neural network model are initialized.
- 2) Example Storage: A limited number of previously seen data examples are stored to retain the information learned from them.
- 3) Initial training: The model is trained with the initial available data. During training, gradients are generated to update the model parameters.

- 4) Data collection: New input data is received for further learning. This data can come from the same task or new ones.
- 5) Constrained optimization: A random sample of stored examples from previously seen data is selected. Optimization is performed on the new input data, ensuring that loss in the selected examples does not increase during this process. This is achieved by calculating and applying gradients in the opposite direction to the gradients from the stored examples.
- 6) Parameter update: The model parameters are updated using the gradients calculated in the previous step.
- 7) Performance evaluation: The model's performance is evaluated on new data or a specific metric to determine its performance in the current task.
- 8) Repeat steps 4-7: Steps 4 to 7 are repeated for each new task or batch of data received.

B. CNN Initialization

The first step of the continual learning method (see Section II-I-A) is to initialize a CNN. A previous work [23] was carried out for this purpose. This previous work designed a study in which 35 Alzheimer's patients were monitored by smartphones with accelerometers over a week in a daycare center. The accelerometer used in this study records the acceleration forces experienced by the patient through a smartphone. These forces are measured in $\text{meter/second}^2 (m/s^2)$ and are captured along the three spatial axes. The recorded acceleration forces include the effects of both gravity and the patient's movements, such as walking. This study generated a dataset storing all the data generated by these accelerometers. The patients did not have to perform any specific task during this period of time. They had absolute freedom to carry out their daily activities in the daycare center. A specialist in neuropsychology at the daycare center placed the phones in the patients' pockets, regardless of where they were. The smartphone's accelerometer sensor produces a data sequence for each patient, capturing changes in acceleration across the X, Y, and Z axes over time. The proposed methodology utilizes these three-dimensional data features along the temporal dimension to forecast the stage of AD. Data were collected from a cohort of 35 patients representing varying stages of AD: 7 in the early stage, 18 in the moderate stage, and 10 in the severe stage. These patients moved without any prior instructions, ensuring the absence of initial biases. Modifications in the positioning of the smartphone should not pose an issue, as they arise from the patient's activities and movements, necessitating the system's ability to adapt by learning from these changes. The study prioritizes a broader range of data variance, enabling the system to glean more insights without constraining device mobility. This approach also serves to underscore the effectiveness of CLADSI in operating within uncontrolled environments.

Regarding the daily activities, it is important to note that no specific instructions were given to the patients, and their activities were not annotated. The intention was to mimic real-life scenarios, allowing for experimentation in an environment that replicates natural behavior. It's important to note that the periods

TABLE II
DISTRIBUTION OF DATA OBTAINED FROM MONITORED PATIENTS

Stage	Number of patients	Samples
Early	7	41
Middle	18	100
Late	10	46

when patients wore the accelerometer did not coincide with their sleep periods.

The dataset also includes information on the stage of the disease for each patient. Thus, each patient was assigned one of the following labels in the dataset to describe the state of the pathology: (i) early, (ii) middle, (iii) late. These labels are based on a previous diagnosis of a neurology using the Global Deterioration Scale (GDS) [24] that uses the following 7 point scale to quantify the pathology stage:

- 1) No cognitive decline.
- 2) Very mild cognitive decline.
- 3) Mild cognitive decline.
- 4) Moderate cognitive decline.
- 5) Moderately severe cognitive decline.
- 6) Severe cognitive decline.
- 7) Very severe cognitive decline.

The correspondence between the dataset labels and the GDS values is as follows: *early stage* (7 patients) includes 2 and 3 levels on the GDS; *middle stage* (18 patients) includes 3, 4 and 5 levels on GDS; *late stage* (10 patients) includes 6 and 7 levels on GDS. Patients with a GDS of 1 were excluded from the study as they were considered healthy. However, the study includes patients with other GDS levels, based on studies that examined patients' gait (see Section II-B). These studies revealed minor variations in gait that may correspond to AD and mild cognitive impairment (levels 2 and 3 of the GDS) and do not significantly impact daily life. Therefore, patients with any GDS level from 2 onwards were included in the study.

After a week, 187 samples were obtained from the different patients. The dataset was unbalanced, with more data from the middle stage than from the other two stages. A summary of the data obtained is shown in Table II.

Commercial smartphones with their accelerometers were used to conduct experiments replicating real-life scenarios, employing devices commonly accessible to individuals with AD. However, it is important to acknowledge that these accelerometers may encounter issues and irregularities, as they are not as advanced as the devices used in specialized movement study laboratories that are directly attached to the body. Thus, the data obtained were of different lengths and had some irregularities, as the frequency of data collection was not stable. As a result, there were some time jumps between consecutive measurements. To address these potential irregularities, a sufficiently wide time window was adopted to ensure a representative sample of the data. This means that even if the commercial mobile phones showed fluctuations in the number of values provided by their accelerometers, enough information was gathered within each window to facilitate proper analysis. It is noteworthy that due to potential variations in the number of samples obtained from

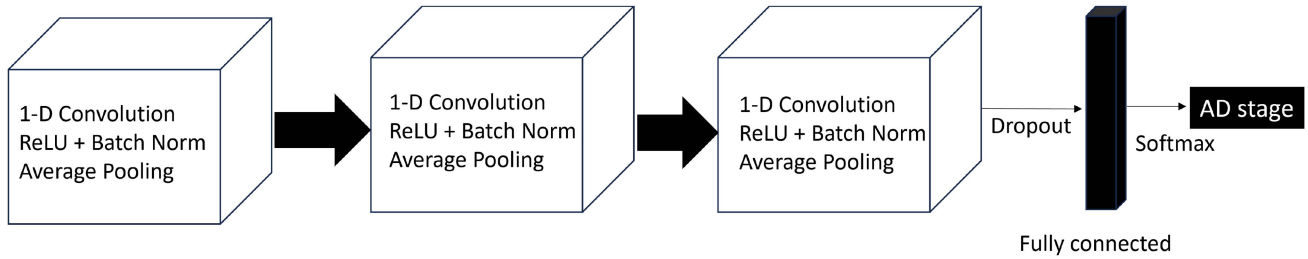


Fig. 1. CNN architecture.

TABLE III

ARCHITECTURE OF THE NETWORK USED TO PROCESS THE DATA SEQUENCES, SHOWING THE DIFFERENT LAYERS, THE HYPERPARAMETERS AND THE WEIGHTS TO BE TRAINED

Layer	Hyper-parameters	Output shape	Number of weights
1D-Convolution	50 filters, size 8	$10,797 \times 50$	1250
ReLU	-	$10,797 \times 50$	-
Batch Normalization	-	$10,797 \times 50$	200
Average Pooling	size 5	$2,159 \times 50$	-
1D-Convolution	100 filters, size 16	$2,144 \times 100$	80,100
ReLU	-	$2,144 \times 100$	-
Batch Normalization	-	$2,144 \times 100$	400
Average Pooling	size 10	214×100	-
1D-Convolution	200 filters, size 32	183×200	640,200
ReLU	-	183×200	-
Batch Normalization	-	183×200	800
Average Pooling	size 10	18×200	-
Flatten	-	3600	-
Dropout	probability 0.75	3600	-
Fully Connected	500 neurons	500	1,800,500
ReLU	-	500	-
Fully Connected	3 neurons	3	1,503
Softmax	-	3	-

different patients, a wide window was chosen to compensate for these disparities and ensure the robustness of the obtained results.

To homogenise the data, a two-stage preprocessing step was conducted. First, the data were grouped every 0.1 seconds, obtaining the mean of each one of them to represent them, thus removing irregularities and avoiding outliers and errors in the measurements. After that, taking into account that the sequences were approximately 1 h long, it was decided to divide each of them into 5 parts, thus making data augmentation with it and obtaining 935 sequences as a result. It is important to note that data from the same patient should not fall into both train and test splits, since the process of training would otherwise be flawed.

Finally, as the CNN needs all inputs to have the same shape, it was decided to take the longest data (10,804) and for this to be set as the input length. Shortest data were 0-extended to that length. Future new data longer than that will be cut down to that length. Alternatively, it could be split into several parts of similar length and then filled with 0's.

The architecture of the resulting CNN contains several sequential layers (see Fig. 1), including three blocks with 1D-Convolutions, ReLU activations and pooling layers; and finally, two fully connected layers to give a final prediction. This prediction is a vector of length 3, each one of them representing the probability of the disease being at an early, middle or late stage, all of them adding up to 1.

The different layers, their output shapes and the weights of each of them are described in Table III. The network has a total of 2,524,253 trainable parameters and 700 non-trainable ones, the latter coming from the Batch normalization layers. This network with the described configuration was the one that obtained the best results. The primary objective was to compare CLADSI (continual learning paradigm) with a CNN that achieved the best success rate with the dataset using a non-continual learning paradigm. While other architectures have shown promising results in accelerometer data analysis, they were trained on different datasets [25], [26], [27]. This approach allowed us to assess the efficacy of continual learning techniques in the context of our specific problem domain. The motivation behind the selection of network hyperparameters, such as the number of epochs, batch size, learning rate, and the number of random training instances selected, was to take as a reference the CNN that had the highest success rate, providing a benchmark for comparison with CLADSI.

C. Model Evaluation and Training

In order to simulate a continual learning environment and perform steps 2 to 8 of the method (see Section III-A), the acquired dataset has been divided into several parts or which are passed to the system in different experiences. The term experience is used to denote a dataset presented to the learning system

for training or evaluation. For this case, three environments have been simulated, each with 2, 3, and 4 data experiences, respectively. After dividing into train and test splits (80% and 20% each), training is divided in the given parts.

For the training process to be carried out, the selected loss function is multi-class cross-entropy, a commonly used function for solving classification problems. It is defined as:

$$L(y, \hat{y}) = - \sum_{k=1}^N y_k * \log(\hat{y}_k) \quad (1)$$

where y is the one-hot encoding vector of the true class, \hat{y} is the vector of predictions, and N is the number of classes.

In order to evaluate the model, several metrics have been selected, four of them already used in previous work and one of them used to evaluate the efficiency of the continual learning method. The two main metrics used are accuracy and F1-score. Accuracy is defined as:

$$Acc = \frac{\#Correct\ predictions}{\#All\ examples} \quad (2)$$

The F1-score and the two secondary metrics from which it is calculated, precision and accuracy, are defined as:

$$Prec = \frac{TP}{TP + FP} \quad Rec = \frac{TP}{TP + FN}$$

$$F1 - score = 2 * \frac{Prec * Rec}{Prec + Rec} \quad \text{or}$$

$$F1 - score = \frac{2 * TP}{2 * TP + FN + FP} \quad (3)$$

where TP, FN and FP stand for True Positives, False Negatives and False Positives respectively.

Since it is a multi-class problem, TP, FP and FN cannot be calculated normally. To do this, a calculation should be made individually for each class, one against the others, and then calculating a single one. In this case, preference is given to the weighted average for combining the metrics, taking into consideration the number of examples in each class.

To evaluate the performance of the selected continual learning method, the forgetting measure [28] has been used. This calculates how much information from previous tasks or experiences is forgotten in the process of training new tasks. The lower the forgetting, the better the continual learning method in maintaining memory. It is important to note that if the accuracy (or F1-score) is low, there is not much information to forget, so other metrics should be considered when using this one. The amount of forgetting in task i is calculated as follows:

$$F_i = \frac{1}{i-1} \sum_{j=1}^{i-1} f_{i,j}$$

$$\text{where } f_{k,j} = \max_{l \in \{1, \dots, k-1\}} acc_{l,j} - acc_{k,j}, \forall j < k \quad (4)$$

where j and l are iteration variables and k is the total number of tasks. This metric is calculated at the end of the continual learning setting, obtaining a value that represents how much information from the first tasks the network has forgotten.

TABLE IV
RESULTS OBTAINED WITH THE DIFFERENT SPLITS OF THE PROPOSED METHOD

Experiences	Accuracy	F1-Score	Forgetting
2	86.94% (± 1.36)	86.60% (± 1.47)	6.89 % (± 1.16)
3	86.48% (± 0.67)	86.11% (± 1.12)	8.80 % (± 1.37)
4	84.37% (± 0.88)	84.04% (± 1.22)	9.47 % (± 1.65)

In order to obtain a better and more homogeneous evaluation of the proposed system, k-fold validation has been used, dividing the dataset into 10 parts and performing 10 different training processes. To do this, and so as to maintain the 80%–20% ratio in training-testing, two of the splits are taken for testing and the rest for training. Taking that into account, there were 748 examples for training and 187 for testing. For each one of the continual learning experiences, there were 374 examples of each experience for the 2-experience setting, 249 examples for the 3-experience setting; and 187 examples for the 4-experience setting.

The system was trained in all cases for 100 epochs per experience, with a batch of 32 and a learning rate of $5 * 10^{-5}$. The memory available to the algorithm was 64 examples in total, and the number of random examples selected to conduct the training of new tasks was 32.

IV. RESULTS

After implementing CLADSI as described in Section III, three trials were carried out, the continual learning environment set by splitting up the training dataset, in each case, into 2, 3 and 4 parts respectively. After executing 10 folds for each trial using the k-fold strategy, the metrics (accuracy, F1-score and forgetting) were obtained and, the metrics aggregated with their average and their standard deviation, as shown in Table IV.

From Table IV it can be extracted that the continual learning setting obtains promising results, but there is a large impact due to the size of the dataset entered in each case, with this becoming more noticeable the more experiences it contains. Moreover, the forgetting metric, calculated over the training sets of previous tasks (all except the last one) shows that there is a slight forgetfulness of information already learnt.

Fig. 2(a)–(c) show the best fold for each of the trials. It can be seen from these graphs that, when switching tasks, there is a slight forgetting of previous tasks, although the test set improves as new tasks are visited, as it is fed with a variety of information from all sets.

On the other hand, it is worth noting that, although Table IV shows the forgetting rate is greater the more experiences there are in the training set, it can be seen in the graph that this is due to the fact that in these cases higher accuracies are achieved in each training task. This is attributed to the limited amount of data available. Given the use of a deep learning model, it tends to overfit on this small dataset. Hence, it can be inferred that the continual learning system also aids in reducing overfitting by stabilizing knowledge as new data is entered. As has been observed, the system achieves the goal of continuously improving as it receives data, although it is recommended to wait for a

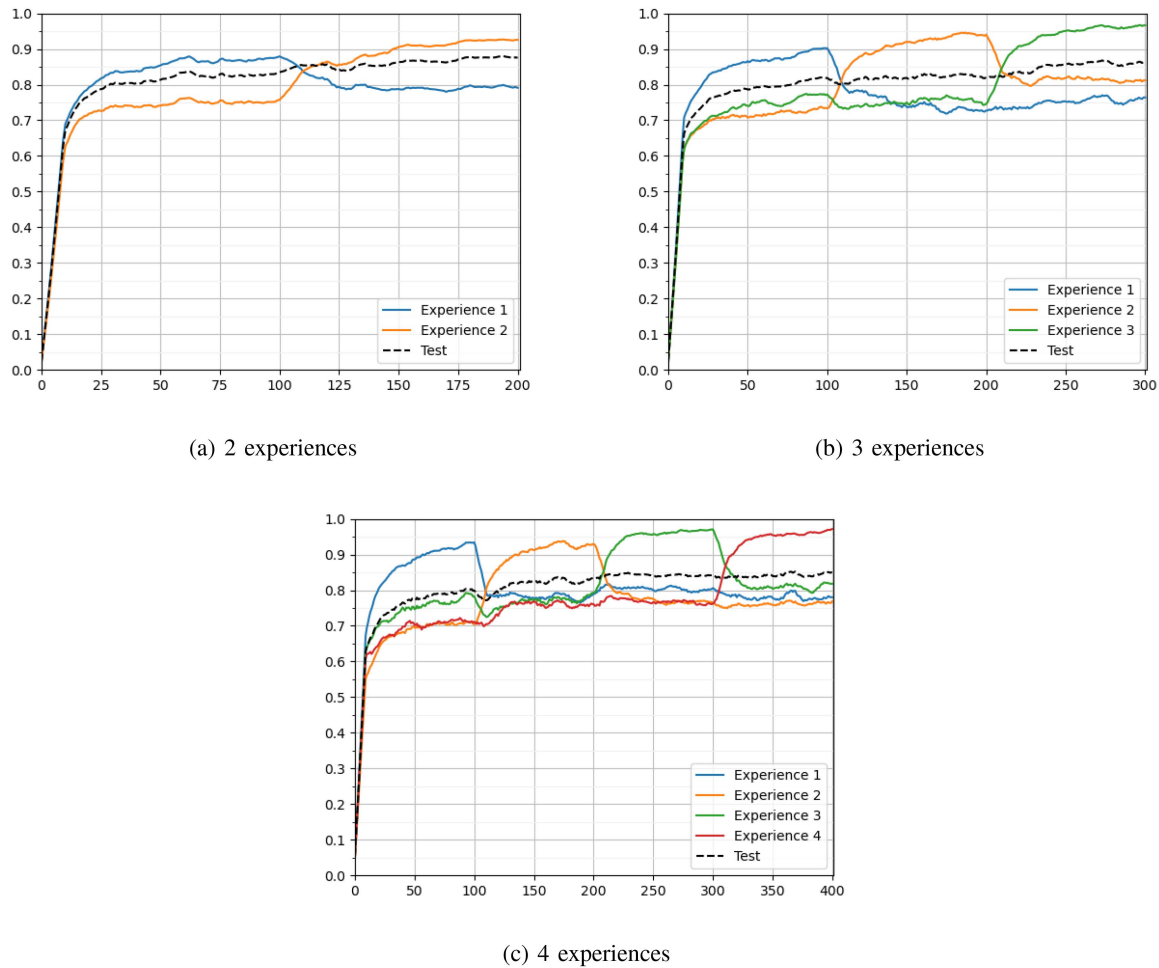


Fig. 2. Plots of the accuracy of the different trials on training over the different tasks, in colored lines, and test, in black dashed line. Tasks change every 100 epochs

TABLE V

COMPARISON OF THE RESULTS OBTAINED IN THE DIFFERENT TESTS CARRIED OUT WITH THE DATASET DESCRIBED

Model	Train sequences	Precision				Recall				F-Score	Acc.
		Early	Middle	Late	Weight.Avg	Early	Middle	Late	Weight.Avg		
Adaptive Boosting (AB)	748	0.51	0.85	0.72	0.76	0.73	0.85	0.50	0.74	0.74	74%
Decision Tree (DT)		0.68	0.84	0.62	0.76	0.70	0.84	0.61	0.76	0.76	76%
K-Nearest-Neighbor (KNN)		0.85	0.83	0.80	0.83	0.70	0.93	0.70	0.83	0.82	83%
Logistic Regression (LR)		0.55	0.76	0.68	0.70	0.36	0.89	0.57	0.72	0.70	72%
MultiLayer Perceptron (MLP)		0.76	0.84	0.77	0.81	0.76	0.90	0.65	0.81	0.81	81%
Random Forest (RF)		0.87	0.88	0.89	0.88	0.79	0.99	0.67	0.88	0.87	88%
Support Vector Machine (SVM)		0.66	0.86	0.84	0.82	0.76	0.94	0.57	0.81	0.81	81%
Convolutional Neural Network (CNN)		0.88	0.92	0.90	0.91	0.86	0.94	0.86	0.90	0.90	91%
CLADSI (CNN 2 exp.)		374	0.8	0.91	0.81	0.86	0.91	0.87	0.84	0.87	0.87
CLADSI (CNN 3 exp.)	249	0.88	0.9	0.76	0.86	0.81	0.91	0.82	0.87	0.86	86%
CLADSI (CNN 4 exp.)	187	0.87	0.87	0.76	0.84	0.76	0.89	0.82	0.84	0.84	84%

The size of the training set used in each case is also shown.

certain volume of data to be available before training the system again.

Table V shows a comparison with previously tested models [29]. In the evaluation of other models used in the experiment (such as Adaptive Boosting, Decision Tree, K-Nearest-Neighbor, Logistic Regression, MultiLayer Perceptron, Random

Forest, Support Vector Machine), configurations yielding the highest success rates were selected. This ensures a more appropriate comparison between CLADSI and these techniques. It can be clearly seen that CLADSI obtains excellent results, considering that the data is very divided, that in each test very little data is given to train in each experience (374, 249 and

187) added to the difficulty of maintaining memory of previous given examples. The continual learning paradigm offers several distinct advantages over traditional methods.

Despite processing significantly less data compared to the traditional CNN [23], CLADSI achieves a comparable success rate (see Table V). This underscores the efficiency of CLADSI in delivering similar results while utilizing a smaller dataset. These findings highlight the effectiveness of CLADSI in managing data streams effectively for AD classification tasks.

One of the key advantages of the proposed method is its ability to continuously learn and adapt to new data over time, thus enhancing adaptability. Traditional approaches often rely on fixed models that may become outdated as new data becomes available. Moreover, by leveraging continual learning, the method can efficiently incorporate new information, leading to significant time and resource savings compared to traditional approaches. This efficiency is particularly valuable as it avoids the need for periodic updates or retraining of the model.

Additionally, the continual learning paradigm enables this method to provide real-time predictions or insights, making it particularly well-suited for applications where timely decision-making is critical. In contrast, traditional methods may struggle to keep pace with rapidly changing data streams or may require batch processing, which can mean there are delays in decision-making.

However, it is important to acknowledge that the proposed method also has some limitations. While continual learning allows the method to adapt to new data, it typically requires an initial training phase to establish a baseline model. This initial training phase may require a sufficient amount of labeled data and computational resources, which could be a limitation in scenarios where labeled data is scarce or expensive to obtain. Moreover, the results of the continual learning paradigm show slightly lower accuracy rates in some cases. Lastly, in clinical settings—especially when the software serves as a medical device and is certified by regulatory bodies like the FDA or EMA for use in patient therapy—continuing to train, tune, or adjust the network weights is not feasible.

V. DISCUSSION

The continual learning process generates a model that achieves an accuracy of 86.94%, 86.48% and 84.37% for 2, 3 and 4 experiences respectively. Previous experimentation with this dataset applying a non-continual learning process achieved an accuracy of 90.91% [23]. Although this previous experimentation reports higher accuracy rates compared to CLADSI, it is important to consider the underlying differences in their methodologies. The previous experimentation utilized a traditional approach with a CNN model, while CLADSI introduces the novel concept of continual learning algorithms and focuses on analyzing data from a continuous stream of motion sensors, allowing the CNN to self-configure as it receives new data.

While CLADSI achieves slightly lower accuracy rates, it offers distinct advantages that contribute to its overall value. In contrast to traditional methods that often require retraining or manual adjustments, CLADSI employs an A-GEM based

approach that preserves previously learned information from past tasks. This method continuously self-adjusts based on new data, allowing it to adapt dynamically to changes in a patient's condition or surroundings. This capability has profound implications in medical scenarios, particularly where AD is concerned. Firstly, it enables early detection and monitoring of gait disturbances, which can signal the onset of AD. By continually analyzing gait patterns in real-time, CLADSI can offer timely insights into disease progression and potentially assist in early intervention efforts. Furthermore, the ability to update the model autonomously without human intervention boosts the efficiency and scalability of medical monitoring systems. Healthcare professionals can rely on automated algorithms to continuously monitor patients, freeing up time and resources for other critical tasks. Additionally, the self-adjusting nature of the A-GEM method allows for personalized and dynamic modifications to the analysis process, enhancing the accuracy and relevance of diagnostic information.

The proposed method in CLADSI exhibits promising potential in the field of deep learning solutions for healthcare, offering an efficient and self-configuring approach to analyzing data from motion sensors. However, as can be seen in Table I, recent works in the literature have not considered a methodology that automatically adapts the deep learning model to this continuous streaming of data. Limitations of these works include static model configurations, lack of adaptability to changes in patient behavior, and the need for manual intervention for updates. In contrast, CLADSI addresses these limitations by introducing a continual learning framework allowing the CNN to automatically adapt to new data, thereby enhancing its ability to efficiently identify AD stages without manual intervention.

VI. THREATS TO VALIDITY

While this study presents promising findings regarding the feasibility of applying the continual learning paradigm to identify the state of AD using accelerometer data, it is important to recognize and address the following potential threats to the validity of the experimentation:

- **Sample size:** One potential threat to the validity of this study is the relatively small sample size, consisting of only 35 Alzheimer's patients. While a larger sample size would enhance generalizability, this does not diminish the validity of the findings. The purpose of this study was to explore the feasibility of using continual learning algorithms with motion sensor data in Alzheimer's patients, and the results provide valuable insights and promising outcomes. Future research can certainly consider expanding the sample size to further validate and extend the findings. One approach could be to replicate the experiment in another daycare center with a different cohort of Alzheimer's patients. This would help assess the generalizability of the findings across different populations.
- **Lower success rate in continual learning paradigm:** The lower success rate in the continual learning paradigm, attributed to the lesser availability of information in each auto-configuration of the model, is an implicit limitation.

However, this does not invalidate the proposed method. It highlights a challenge that could be addressed through further research and optimization. By exploring techniques such as hyperparameter tuning, regularization, or architectural modifications, it is possible to improve the success rate and achieve better performance in continual learning scenarios.

- **Cross-dataset scenario:** The question of whether the proposed methods can be applied in a cross-dataset scenario is worth considering. While this study focused on a specific dataset of Alzheimer's patients monitored in a daycare center, the potential applicability of the proposed method in cross-dataset scenarios has not been ruled out. The discussion of generalization across datasets highlights the need for further investigation and validation using external datasets. Applying techniques such as transfer learning or domain adaptation could facilitate the adaptation of the trained model to new datasets, increasing its versatility and robustness.
- **Medical relevance and practical use:** It is acknowledged that predicting or diagnosing AD based solely on motion sensor data is not medically relevant, as physicians typically rely on multimodal data for personalized and accurate decision-making. However, this limitation does not undermine the proposed method but rather positions it as a complementary tool in the medical domain. The proposed approach can aid in the continuous monitoring and early detection of the disease, potentially enhancing existing diagnostic practices. Further research should explore the integration of the proposed method with other clinical data modalities to improve diagnostic accuracy and clinical relevance.

In conclusion, while several threats to validity are present in this study, they do not invalidate the proposed method. Instead, they highlight areas for further research and improvement. The findings of this study provide valuable insights into the application of continual learning algorithms with motion sensor data for AD identification. In future research, these threats can be mitigated through several approaches. Firstly, increasing the sample size would enhance statistical power and allow for more robust conclusions. Additionally, conducting longitudinal studies with larger and more diverse cohorts could provide a better understanding of the model's performance across different populations. Moreover, implementing cross-validation techniques and testing the model on independent datasets can help validate its generalizability. Furthermore, refining the continual learning paradigm by exploring alternative algorithms or incorporating mechanisms for model reinitialization or regularization could potentially improve success rates.

VII. CONCLUSION

In this work, a novel approach has been presented for identifying the stage of AD using continual learning algorithms applied to motion sensor data. The method leverages the power of deep learning and the availability of wearable sensors, specifically smartphone accelerometers, to enable continuous monitoring and accurate classification of AD stages.

By employing a continual learning paradigm, the approach addresses the challenge of auto-configuring deep learning models in the context of a continuous stream of sensor data. The results obtained demonstrate the feasibility and effectiveness of the method, achieving high accuracy rates of 86.94%, 86.48%, and 84.37% for 2, 3 and 4 experiences, respectively.

One of the key advantages of this approach is its ability to identify the stage of AD, providing valuable insights into disease progression and facilitating personalized treatment and intervention strategies. This has significant implications for both patients and healthcare providers, enabling more targeted and timely interventions to improve patient outcomes.

Furthermore, the approach offers several additional benefits. By using continual learning algorithms, the models can adapt and self-configure in real-time as new data streams in, reducing the need for manual intervention and saving time and computational resources. This efficiency and responsiveness means the method is particularly suitable for healthcare systems that rely on continuous monitoring of patients using wearable sensors. To begin with, a CNN is trained from a random point with all the data initially available. In a medical environment, it is usually difficult to obtain data, so presumably a small volume will be available at start. Then, using the A-GEM algorithm, the system will learn from any new given data, taking advantage of its continual learning capabilities. This makes it possible to take advantage of the intelligent system from the start, deploy it quickly and improve it over time.

Moreover, the approach leverages widely accessible sensors such as smartphone accelerometers, meaning it can be easily integrated into existing healthcare systems. This opens up opportunities for long-term, remote monitoring of patients, providing a more comprehensive understanding of their daily activities and disease progression. The continuous nature of the monitoring approach offers a more naturalistic assessment of patients' routines and allows for early detection of changes or deterioration in their condition.

In future research, the aim is to further refine the proposed approach for continuous monitoring and early detection of AD, recognizing the critical importance of closely tracking disease progression, especially between clinical visits for early-stage patients. Remote monitoring can be seen as a promising avenue for continuous assessment, empowering clinicians to make informed decisions based on real-time data insights. To better support this strategy, the plan is to enhance the method to approximate the GDS at a finer level. This refinement will offer clinicians a more comprehensive understanding of their patients' disease status, complemented by two additional approaches. Firstly, we will develop explainability mechanisms to provide descriptive information on variations in patients' mobility patterns associated with AD progression. Secondly, experiments will be conducted using data on patients' mobility in home environments, allowing us to study less controlled activities compared to those in day centers. By combining these efforts, the objective is to provide clinicians with detailed insights into patients' conditions, enabling more informed decisions regarding medication and therapies, and ultimately enhancing patient care outcomes.

REFERENCES

- [1] J. Thunell et al., "Drug therapies for chronic conditions and risk of Alzheimer's disease and related dementias: A scoping review," *Alzheimer's Dement.*, vol. 17, no. 1, pp. 41–48, 2021.
- [2] B. Dubois et al., "Clinical diagnosis of Alzheimer's disease: Recommendations of the international working group," *Lancet Neurol.*, vol. 20, no. 6, pp. 484–496, 2021.
- [3] Y. Cedervall, K. Halvorsen, and A. C. Åberg, "A longitudinal study of gait function and characteristics of gait disturbance in individuals with Alzheimer's disease," *Gait Posture*, vol. 39, no. 4, pp. 1022–1027, 2014.
- [4] V. R. Varma et al., "Continuous gait monitoring discriminates community-dwelling mild Alzheimer's disease from cognitively normal controls," *Alzheimer's Dement.: Transl. Res. Clin. Interv.*, vol. 7, no. 1, 2021, Art. no. e12131.
- [5] Y.-L. Hsu et al., "Gait and balance analysis for patients with Alzheimer's disease using an inertial-sensor-based wearable instrument," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1822–1830, Nov. 2014.
- [6] R. Baker, "Gait analysis methods in rehabilitation," *J. Neuroengineering Rehabil.*, vol. 3, no. 1, pp. 1–10, 2006.
- [7] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Netw.*, vol. 113, pp. 54–71, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608019300231>
- [8] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," *Psychol. Learn. Motivation*, vol. 24, pp. 109–165, 1989.
- [9] I. J. Goodfellow, M. Mirza, D. Xiao, A. C. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," in *Proc. 2nd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., Apr. 14–16, 2014. [Online]. Available: <http://arxiv.org/abs/1312.6211>
- [10] M. D. Lange et al., "A continual learning survey: Defying forgetting in classification tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3366–3385, Jul. 2022.
- [11] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5533–5542.
- [12] R. Aljundi et al., "Online continual learning with maximally interfered retrieval," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2019, Art. no. 1063.
- [13] O. Beauchet et al., "Poor gait performance and prediction of dementia: Results from a meta-analysis," *J. Amer. Med. Directors Assoc.*, vol. 17, no. 6, pp. 482–490, 2016.
- [14] R. M. Ardle, B. Galna, P. Donaghy, A. Thomas, and L. Rochester, "Do Alzheimer's and lewy body disease have discrete pathological signatures of gait?," *Alzheimer's Dement.*, vol. 15, no. 10, pp. 1367–1377, 2019.
- [15] H. Zhou et al., "Digital biomarkers of cognitive frailty: The value of detailed gait assessment beyond gait speed," *Gerontology*, vol. 68, no. 2, pp. 224–233, 2022.
- [16] A. S. Alharthi, S. U. Yunas, and K. B. Ozanyan, "Deep learning for monitoring of human gait: A review," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9575–9591, Nov. 2019.
- [17] M. Seifollahi et al., "Detection of mild cognitive impairment from gait using adaptive neuro-fuzzy inference system," *Biomed. Signal Process. Control*, vol. 71, 2022, Art. no. 103195. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809421007928>
- [18] B. Ghoraani, L. N. Boettcher, M. D. Hssayeni, A. Rosenfeld, M. I. Tolea, and J. E. Galvin, "Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning," *Biomed. Signal Process. Control*, vol. 64, 2021, Art. no. 102249.
- [19] Z. You, Z. You, Y. Li, S. Zhao, H. Ren, and X. Hu, "Alzheimer's disease distinction based on gait feature analysis," in *Proc. IEEE Int. Conf. E-Health Netw., Appl. Serv.*, 2021, pp. 1–6.
- [20] S. Gillain et al., "Gait speed or gait variability, which one to use as a marker of risk to develop alzheimer disease? A pilot study," *Aging Clin. Exp. Res.*, vol. 28, no. 2, pp. 249–255, 2016.
- [21] J. F. Pedrero-Sánchez, J.-M. Belda-Lois, P. Serra-Añó, M. Inglés, and J. López-Pascual, "Classification of healthy, alzheimer and parkinson populations with a multi-branch neural network," *Biomed. Signal Process. Control*, vol. 75, 2022, Art. no. 103617.
- [22] A. Chaudhry, M. Ranzato, M. Rohrbach, and M. Elhoseiny, "Efficient lifelong learning with a-gem," 2019. [Online]. Available: https://openreview.net/forum?id=Hkf2%5C_sCS5FX
- [23] S. Bringas, S. Salomón, R. Duque, C. Lage, and J. L. Montaña, "Alzheimer's disease stage identification using deep learning models," *J. Biomed. Inform.*, vol. 109, 2020, Art. no. 103514.
- [24] C. M. Belden, C. Burciu, and M. Sabbagh, "Global Deterioration Scale (Individuals)," in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed., Dordrecht, The Netherlands: Springer, 2014, pp. 2562–2564, doi: [10.1007/978-94-007-0753-5_1173](https://doi.org/10.1007/978-94-007-0753-5_1173).
- [25] A. Papadopoulos, K. Kyritsis, L. Klingelhoefer, S. Bostanjopoulou, K. R. Chaudhuri, and A. Delopoulos, "Detecting parkinsonian tremor from IMU data collected in-the-wild using deep multiple-instance learning," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2559–2569, Sep. 2020.
- [26] A. Kirmizis, K. Kyritsis, and A. Delopoulos, "A bottom-up method towards the automatic and objective monitoring of smoking behavior in-the-wild using wrist-mounted inertial sensors," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2021, pp. 6867–6870.
- [27] K. Kyritsis, C. Diou, and A. Delopoulos, "A data driven end-to-end approach for in-the-wild monitoring of eating behavior using smart-watches," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 1, pp. 22–34, Jan. 2021.
- [28] A. Chaudhry, P. K. Dokania, T. Ajanthan, and P. H. Torr, "Riemannian walk for incremental learning: Understanding forgetting and intransigence," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 532–547.
- [29] A. Nieto-Reyes, R. Duque, J. L. Montaña, and C. Lage, "Classification of Alzheimer's patients through ubiquitous computing," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1679.