STEP-UP: Enabling low-cost IMU sensors to predict the type of dementia during everyday stair climbing

Catherine Holloway[1, 2, 3*], William Bhot[2, 3], Keir Yong[4], Ian McCarthy[5], Tatsuto Suzuki[5], Amelia Carton[4], Biao Yang[5], Robin Serougne[5], Derrick Boampong[5], Nick Tyler[6], Sebastian J Crutch[4], Nadia Berthouze[1], and Youngjun Cho[1,2,3]

[1]UCL Interaction Centre, University College London, London, United Kingdom

[2]Global Disability Innovation Hub, University College London, London, United Kingdom

[3]Department of Computer Science, UCL, London, United Kingdom

[4]Dementia Research Centre, Institute of Neurology, University College London, London, United Kingdom

[5]Pedestrian Accessibility and Movement Environment Laboratory, Department of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom

[6]Universal Composition Laboratory, University College London, London, United Kingdom

*** Correspondence:**
Catherine Holloway
c.holloway@ucl.ac.uk

Keywords: Health – Clinical[1], Wearable Computers[2], Empirical study that tells us about people[3], Lab Study[4], Dementia[5]

Abstract

Posterior Cortical Atrophy is a rare but significant form of dementia which affects peoples' visual ability before their memory. This is often misdiagnosed as an eyesight rather than brain sight problem. This paper aims to address the frequent, initial misdiagnosis of this disease as a vision problem through the use of an intelligent, cost-effective, wearable system, alongside diagnosis of the more typical Alzheimer's Disease. We propose low-level features constructed from the IMU data gathered from 35 participants, while they performed a stair climbing and descending task in a real-world simulated environment. We demonstrate that with these features the machine learning models predict dementia with 87.02% accuracy. Furthermore, we investigate how system parameters, such as number of sensors, affect the prediction accuracy. This lays the groundwork for a simple clinical test to enable detection of dementia which can be carried out in the wild.

Introduction

The rate of people living with dementia is increasing. Alzheimer's Disease (AD) is the most common cause of dementia and is often seen as simply part of the ageing process and something which will affect most people (International 2019) as the average living age increases. AD is a progressive disease which affects a person's memory and therefore their ability to conduct activities of daily

living independently which decreases their quality of life (Gale, Acar, and Daffner 2018). However, AD is not a single disease type, instead there is the typical presentation and a number of atypical presentations (Graff-Radford et al. 2021). Posterior Cortical Atrophy (PCA) is one such atypical presentation which typically results in 'a progressive, often striking, and fairly selective decline in visual-processing skills and other functions that depend on the parietal, occipital, and occipitotemporal regions of the brain' (Crutch et al. 2012). Different types of AD may often be misdiagnosed until quite advanced. This is indeed the case for PCA where the atypical vision-based symptoms present themselves at an early age (typically emerging during 50-65 years old) leading to a simple vision-problem diagnosis (Crutch et al. 2012). Therefore, it is important to develop methods that can identify AD regardless of its type so that people with rare forms can efficiently get the treatment they need. We do this by building on previous studies into everyday walking tasks detection.

People with typical Alzheimer's Disease (tAD) have characteristic issues when navigating their everyday environments (I. McCarthy et al. 2019) with a noticeable general decline in gait patterns (Valkanova and Ebmeier 2017). Previous lab-based research has demonstrated differences in gait parameters such as step-time and walking speed between people with dementia and age-matched controls (Cedervall, Halvorsen, and Åberg 2014; Marquis et al. 2002; Rosso et al. 2017; Verghese et al. 2007; Waite et al. 2005; L. Wang et al. 2006). These studies indicate that the decline is linked to both phenotype and stage of the disease (Allali et al. 2016; del Campo et al. 2016; Castrillo et al. 2016; I. McCarthy et al. 2019; Yong et al. 2020). Furthermore, a noticeable decline in gait is thought to predate other cognitive decline (Hall et al. 2000). Therefore, a decline in gait appears to be an appropriate biomarker for the detection of dementia (Montero-Odasso 2016). However, it is important to move out of the laboratory setting to in-the-wild settings for clinical tools to better aid persons with disability (Holloway and Dawes 2016). In the recent disability interactions manifesto (Holloway 2019) the need for in-the-wild data collection was clearly stated. Such data sets were deemed essential to ensure future technologies to aid persons with disabilities such as dementia in living more independently.

This work is part of a wider investigation of gait and spatial navigation in people with dementia in a living lab environment, which specifically focusses on both people with tAD and PCA. Within the field of dementia there is a need for research in living labs, which move beyond highly controlled lab-based settings (Schneider and Goldberg 2020; Duff 2020). The living labs serve as a stepping-stone to full in-the-wild testing (Alavi, Lalanne, and Rogers 2020). Full in-the-wild testing for dementia could reduce the stress of clinical tests for patients and allow for continuous monitoring of decline. Therefore, in this research we aim to pave the way to in-the-wild detection of dementia by discriminating people with dementia from controls in a living lab. Furthermore, we include a rare form of dementia – PCA – that is often missed by clinicians, demonstrating the benefits of this approach to dementia detection. The evidence-based discrimination of dementia, particularly its atypical presentations, not only has clinical applications, but also addresses a key desire of health and social-care professionals for better understanding of rarer presentations of dementia, for appropriate evidence-based assessment (McIntyre et al. 2019). Our apparatus uses low-cost, unobtrusive devices to discriminate dementia, which not only increases the applicability of our research, but also has not been achieved before. Furthermore, we analyse system parameters that led to accurate discrimination, which could aid future research seeking to extend this research or deploy it in the wild.

Therefore, in this paper we focus specifically on the question – can wearable, low-cost, unobtrusive devices be used to detect AD regardless of its presentation? In answering this question, we contribute the following:

- Demonstrate the feasibility of discriminating controls from people with two types of dementia (the more typical Alzheimer's disease (tAD) and a rare form of dementia – Posterior Cortical Atrophy (PCA)) in a simulated real-world environment – a staircase. To do this we analysed data from a low-cost, IMU system using machine learning classifiers. The developed analysis software tools are available at https://github.com/williambhot/detecting_dementia_stairs.
- Examine different system parameters and the direction of traversal that promote accurate discrimination of dementia.
- Release a data set of IMU data from people with tAD, older adults and people with PCA to foster this work in the research community.
- Discuss use cases for the proposed system.

While the primary aim of this study is to discriminate both the rare PCA and more typical Alzheimer's Disease from healthy controls, we also analyse differences in the detection of these two types of the disease by analysing the performance of a ternary model that seeks to discriminate the two types of dementia from each other as well as from controls.

We believe that this research, could provide a key stepping-stone in enabling potential applications in detecting dementia such as a screening tool for healthcare workers and practitioners, general self-screening and support tool. Nevertheless, further research would be required before this is possible to address some of the limitations of this study (such as generalisation issues) and full in-the-wild testing. We discuss this further in the Discussion section.

Related Work

Posterior cortical Atrophy (PCA)

PCA is a rare early-onset syndrome which presents with visual complaints and is most commonly caused by Alzheimer's disease (AD) pathology. PCA has been identified as a distinct clinical syndrome as opposed to just AD with specific, noticeable visual deficits (Mendez, Ghajarania, and Perryman 2002). It also affects literacy, numeracy and gesture (Crutch, Yong, and Shakespeare 2016). People with PCA, as opposed to typical AD (tAD) have better language and memory abilities (Crutch, Yong, and Shakespeare 2016; Firth et al. 2019), but these come at the cost of a greater understanding of the disease and higher levels of depression (Mendez, Ghajarania, and Perryman 2002). Specific interventions need to be developed for people with PCA which help overcome the difficulties they face in visual tasks and help aid better mental health (Mendez, Ghajarania, and Perryman 2002). However, such interventions can only be developed once the disease has been detected and detection is often delayed due to the atypical symptoms compared to tAD and the early onset of the disease (Crutch et al. 2012; Graff-Radford et al. 2021).

Detecting rare forms of dementia like PCA with confidence is not an easy task. People often notice something going wrong with their eyes, e.g. being unable to see a shuttlecock once it has landed on the ground but being able to see it when in flight. The first stop for people following these visual oddities is to visit the optician or GP. It is rare that the symptoms as presented are immediately associated with a form of AD. More generally health and social care practitioners are often unaware of, and find it difficult to appreciate that forms of dementia can affect people's visual abilities (McIntyre et al. 2019).

Dementia Detection

Previous work in the detection of dementia has ranged from mobile-based automatic speech recognition tools (e.g. (Shibata et al. 2018), (Tröger et al. 2018)) to oculomotor performance during web browsing and multimodal interactions with computer avatars (Cano et al. 2017). However, to date these screening tools remain proofs of concept rather than clinical tools.

Previous research has identified that changes in gait are sensitive to dementia, even at early disease stages (Hall et al. 2000), and during the transitional stage between normal cognitive decline and dementia also known as Mild Cognitive Impairment (Schaat et al. 2020; Halloway et al. 2019; Gwak, Woo, and Sarrafzadeh 2018). It was found that a decline in gait predates observable cognitive changes associated with dementia, and gait continues to decline with the progression of dementia (Verghese et al. 2007; Cedervall, Halvorsen, and Åberg 2014; L. Wang et al. 2006; Waite et al. 2005; Marquis et al. 2002). By comparing the gait of healthy age-matched controls to that of people with dementia, clinical research has identified that changes in the pace, rhythm and variability of gait are associated with the decline into dementia (Verghese et al. 2007). Researchers have found people with dementia to have a lower natural walking speed (Verghese et al. 2007; L. Wang et al. 2006; Waite et al. 2005; Marquis et al. 2002), lower cadence, shorter stride length, shorter swing times and longer stance times as well as longer double support times (Verghese et al. 2007). Furthermore, studies have also shown that variability in gait is higher amongst people with dementia, who lack rhythmic and consistent gait (Verghese et al. 2007).

While previous clinical research has helped to identify the changes in gait that occur during the decline into dementia, this research has ignored two important factors that would allow such knowledge to be used for detection of the disease in the wild. Firstly, previous research relies heavily on experiments conducted in laboratory settings that do not mirror the complexities of the real-world environments through which people with dementia must navigate (I. McCarthy et al. 2019). These laboratory experiments usually involve monitoring the gait of participants while they walk along a straight, uninclined path for a short distance and use full biomechanics models to determine changes in gait (Verghese et al. 2007; L. Wang et al. 2006; Waite et al. 2005; Marquis et al. 2002). For example, many use electronic walkways with inbuilt pressure sensors (Verghese et al. 2007; Callisaya et al. 2017; Wittwer, Webster, and Hill 2013) or motion capture systems (Cedervall, Halvorsen, and Åberg 2014). The form factor, complicated setup procedures and price of these measurement systems limit their use in real world environments. Secondly, while some previous studies have analysed different types of dementia (Mc Ardle et al. 2020), previous studies ignore the differences between types of dementia and either focus on one type of dementia (Cedervall, Halvorsen, and Åberg 2014; Callisaya et al. 2017; Wittwer, Webster, and Hill 2013) or consider dementia without looking at its type (L. Wang et al. 2006; Marquis et al. 2002). Furthermore, to our knowledge, gait of people with PCA has only been analysed by previous research in this line of investigation (Carton et al. 2016; I. D. McCarthy et al., n.d.; Ocal et al. 2017; Yong et al. 2020; I. McCarthy et al. 2019; Yong et al. 2018). This research has found that some patients with dementia show a consistent pattern of hesitation (which can be identified from step times) when navigating complex routes (I. McCarthy et al. 2019; Yong et al. 2020; I. McCarthy et al. 2019). However, it was not possible within that task to identify patterns which could be used for predictive purposes. We believe that the regular pattern offered by stairs will help to regularise these irregularities within the gait pattern which would then allow for successful detection of tAD and PCA. Once the feasibility of this approach is established, it will enable a low-cost detection device to be added to footwear. This could enable the detection of dementia in the wild, minimising stressful laboratory tests, and promoting data-driven methods for appropriate detection of dementia for both typical AD and the rarer PCA. Furthermore, the ability of the device to detect the typical Alzheimer's disease (tAD) provides the final product with a much wider number of use cases. The unobtrusive, low-cost nature

of such a device enables its deployment in high-risk populations to continuously monitor changes in risk of developing dementia.

Materials and Methods

In this section, we present the proposed STEP-UP framework and technical details.

Data Collection Protocol

Participants' gait was monitored using Inertial Measurement Units (IMUs) while they climbed a staircase in the living lab environment. This living lab was co-designed by clinical, engineering and computer science researchers, with inputs from patients. The IMUs used were MTw (Xsens Technologies B.V., The Netherlands). They are comprised of an accelerometer, a gyroscope, and a magnetometer (however, the magnetometer was not used for this study). Each participant had a sensor attached to the outside of each heel with the long axis being horizontal, as well as a sensor on the back of the pelvis attached orthogonally to the sensors on the heels (Figure 1). Participants were asked to walk up or down a short flight of stairs consisting of four steps (the dimensions of each step were 23 cm × 112 cm × 25 cm, H × W × D) (Figure 1) in a variety of environmental conditions. These environmental conditions included different lighting levels (low: 20 lux; high: 190 lux) and either the presence or absence of visual cues (i.e. hazard tape over the edge of steps). Each participant was asked to attempt 16 versions of the trial (twice for each combination of conditions – dim light/bright light, visual cues/no visual cues – in the upwards and downwards direction). No constraints were imposed on the way of descending or ascending the stairs. The ordering of trials was randomised for each participant (See Figure 2a).

Participants

Participants were from one of three groups – the group with PCA (containing 11 participants – 6 female and 5 male – of age 64.6 ± 5.9 years, height 168.92 ± 6.49 cm, weight 68.22 ± 13.31 kg, with Mini Mental State Examination (MMSE) score 18.6 ± 6.1), the group with tAD (containing 10 participants – 6 female and 4 male – of age 66.2 ± 5.0 years, height 167.91 ± 11.82 cm, weight 66.21 ± 5.03 kg, with MMSE score 18.6 ± 5.0) and the control group consisting of age matched participants with no diagnosed form of dementia (containing 14 participants – 6 female and 8 male – of age 64.2 ± 4.1 years, height 172.36 ± 13.21 cm, weight 73.23 ± 15.23 kg). The experimental design of having a control group of healthy age-matched participants is the standard experimental protocol used in this field (Callisaya et al. 2017; I. McCarthy et al. 2019). MMSE tests were only conducted on people with dementia, and not on control participants. One-way ANOVAs demonstrated that there were no statistically significant differences between the groups in age ($F(2,32) = 0.506; p = 0.61$), weight ($F(2,30) = 0.404; p = 0.67$) or height ($F(2,31) = 0.580; p = 0.57$). Furthermore, a student t-test showed that there was no difference between MMSE scores for participants in the PCA and tAD conditions ($t(18) = 0; p = 1$). Ethical approval for the study was provided by the National Research Ethics Service Committee London Queen Square, and written informed consent was obtained from all 35 participants.

Pre-processing and Classification Strategy

The data was processed in Python 3.7 (Python Programming Language, RRID:SCR_008394) using standard data processing libraries including NumPy (NumPy, RRID:SCR_008633), SciPy (SciPy, RRID:SCR_008058), Pandas (Pandas, RRID:SCR_018214), Matplotlib (MatPlotLib, RRID:SCR_008624) and Scikit Learn (scikit-learn, RRID:SCR_002577). The data pre-processing

and classification strategy is shown in Figure 2. This process included hyperparameter optimisation on the models to select the best parameters and analysis of how direction of traversal and different system setups affected the performance of this model. This section summarises the methods we used to achieve this. The software tools we developed are released to foster this work in the research community (https://github.com/williambhot/detecting_dementia_stairs).

Exclusion of Participants

On visualising the IMU data – acceleration and gyroscope data – data for some trials was found to be corrupted. Visualising the raw data from these trials showed only noise and no evidence of cyclic, step-like motion (Figure 2b). Therefore, these trials were removed from further analysis.

This resulted in the removal of 11 trials from a total of 527 trials (Table 1). After removing excluded trials, 40.12% of trials were controls, 29.07% were in the PCA condition and 30.81% were in the tAD condition. Up-sampling was conducted on the trials from the different conditions before training any models, so that the models did not overfit to these differences in the frequencies in the groups.

Dead Reckoning and Gait Parameters

Initially we tried to calculate velocity and displacement from the IMU data using a dead-reckoning technique with a zero-offset to account for sensor drift (Park and Suh 2010; Ojeda and Borenstein 2007). Using this we calculated gait parameters that have been previously associated with dementia such as lower walking speed (Verghese et al. 2007; L. Wang et al. 2006; Waite et al. 2005; Marquis et al. 2002) and shorter stride length (Verghese et al. 2007). However, we found that in our current set up it was not possible to conduct dead reckoning with a high enough degree of accuracy for calculating the gait parameters required. We attribute this to the experimental setup as well as issues with controlling the task across participants, especially those with more advanced dementia. See the discussion for more details on this.

Lower-level features

Considering the difficulty of conducting dead-reckoning and calculating gait parameters in a system designed to be useable in the real world, we propose more low-level features that, from a low-cost IMU system, can be more easily designed for real-world use. This involved calculating the vector length of the 3d linear and angular acceleration to obtain the resultant linear and angular acceleration (See Figure 2c):

$$R = \sqrt{x^2 + y^2 + z^2}$$

These two signals – resultant linear acceleration and resultant angular acceleration – were then split into a constant number of windows (k) and the averages of each window (μ_i where i is the number of the window) were used as the features. The windows were calculated in the following way – across the entire dataset, the same number of windows (k) were used and in a single trial these windows were of the same length (l), however, across multiple trials window length was different (See Figure 2c):

$$\mu_i = \frac{\sum_{t=i\times l}^{(i+1)\times l} R_t}{l}$$

Where $i \in [0, k)$ is the number of the current window varying between 0 and $k-1$, k is the total number of windows and t is the current sample for the linear or angular acceleration.

These windowed averages were used as the feature values, allowing a constant number of features for each trial, while providing the model with information from different sections of the trial. The primary reason for using this approach was to have a constant number of features for all trials, which is required by many Machine Learning models. The number of windows was set using hyperparameter optimisation. Specifically, different numbers of windows were experimented with, but it was found that models using a multiple of four windows achieved a higher performance than others and specifically eight windows yielded the best performance (Figure 3). One reason for this could be that there were four steps in the staircase and, therefore, setting the number of windows to a multiple of four provides an approximate way to separate the data based on steps, assuming each step is traversed in approximately the same amount of time in a single trial. However, every participant did not take the same amount of time on each step, and several participants waited for a while on some steps. Therefore, for these participants segmenting the data in this way would not segment the trial by steps. Nevertheless, this was not our motivation for doing this, but rather it was to segment the trial into an equal number of windows so that models that required a fixed number of features could be employed.

Machine Learning Models

We assessed the ability of different machine learning models to classify the data, including decision trees (Random Forest and Gradient Boosting Models) and Multi-Layer Perceptron (MLP) models. To this end, we fit the models to the data and evaluated the models' ability to generalise by testing it on unseen data (see the following section). Furthermore, we chose the parameters of this model through hyper-parameter optimisation discussed later (See Figure 2d).

Two variants of all the models were fit to the data – a binary model to discriminate dementia from control participants and a ternary model to discriminate between controls, tAD and PCA participants. While we were able to discriminate people with dementia from control participants, we were unable to discriminate PCA from tAD with high accuracy (see results section for more details). We suggest that this is because the gait of the two types of dementia was similar to each other and therefore could not be discriminated using these low-level features (see discussion for more details).

Nevertheless, given features ($\mu_i; where\ i \in [0, k)$) the models learnt a mapping (Γ) from features to the probability (p) of this data belonging to the different classes ($c; where\ c = \{control, dementia\}\ or\ c = \{control, PCA, tAD\}$). This is as follows:

$$p(c|\mu_0, \dots, \mu_k) = \Gamma(\mu_0, \dots, \mu_k)$$

Based on the value of this probability for each class, the most likely class for that data can then be ascertained as the class with the maximum probability.

Evaluation of Models

A Leave-One-Person-Out (also called leave-one-subject-out, LOSO) cross validation was used to evaluate the generalization capabilities of our predictions (See Figure 2e). In this method, the model is trained on the data from all but one participant (Cho, Julier, and Bianchi-Berthouze 2019). Predictions are then made on the data from the remaining participant to gauge how well the model performs on unseen data from a participant on which it has not been trained. As data from each model are not independent from one another, the Cochran's Q test was used to determine the significance of the overall accuracy of each model. This was done using the dichotomous 'true' or 'false' prediction for each fold. A pairwise post-hoc Dunn test with Bonferroni adjustments was used

to test for differences between models. All statistical tests were run with a significance level of $\alpha = 0.05$ and were conducted using IBM SPSS V25 (IBM SPSS Statistics, RRID:SCR_019096).

Furthermore, we report accuracy and F1 scores for all models. These are calculated by exhaustively leaving each participant out (as explained above), training the model on the remaining participants and evaluating the model on the participant left out. The accuracy and F1 score were then calculated across all these folds of the data. The accuracy was calculated as the number of correctly classified trials over the total number of trials. F1 scores with respect to each class were calculated as:

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

Hyper-parameter optimization

The hyper-parameters for all models were chosen using hyperparameter optimisation – a standard method in Machine Learning for systematically choosing the parameters of the model that are not directly learnt. All the models were tuned for this study using a type of hyper-parameter tuning – exhaustive grid search (Buitinck et al. 2013) in which variations of the model are run repeatedly using different values of the hyper-parameters, that have been identified manually. The hyper-parameters chosen for the model for the final analyses were the parameters that produced the best performance while conducting the grid search (Table 2). This approach was also used for selecting the number of windows to use in constructing the features (See Figure 3).

Direction of Traversal and System Analysis

A secondary aim of the study was to identify the components of the system that promote a high classification accuracy. This involved analysing: the importance of the three sensors, the importance of the different features and the importance of the direction of traversal of the stairs.

For the analysis of the importance of the sensors, the performance of different variants of the models was analysed. These variants of the models used features from different combinations of the sensors. The importance of the different features was analysed using the tree-based models (ie the Random Forest and Gradient Boosting models), firstly, because they provide methods for determining the importance of features in making a prediction and secondly, due to their high performance. This analysis was done, by calculating the reduction in impurity (or error) that each node (or partition) provides weighted by the probability of reaching that node in the tree and then averaged over all trees to give the final metric of importance. Therefore, importance represents how well the feature portioned the data into the relevant classes weighted by the likelihood of this feature being used in classifying a datapoint. The analysis of traversal direction was done by training the model on all the data, then separating predictions into those made on trials in the upward direction and those made in the downward direction and calculating the accuracy on these subsets separately.

To understand which sensors were most effective a Kruskal-Wallis H-test was conducted and pairwise post-hoc Dunn tests with Bonferroni adjustments were used to determine which sensors to use in further analyses. Finally, a Friedman's Two-Way Analysis of Variance was conducted to understand the importance of features and the influence of upwards and downwards traversal.

Results

Prediction Results

This section presents the results achieved in detecting whether participants had dementia as well as the type of dementia.

In the binary models, trained to discriminate people with dementia from controls, the Random Forest Classifier was the most successful at predicting the presence of dementia, which it accurately did in 87.02% of cases (see Table 3, Figure 4 for more details). Furthermore, the F1 score with respect to control class was 83.14% and 88.38% with respect to the dementia class, both of which were higher than the same for any other model. The Cochran's Q test confirmed the differences between the performance of the models, $\chi2(4, N = 516) = 47.56$, $p < .001$.

In the case of the ternary type-based classification (Control vs tAD vs PCA), the MLP classifier outperforms all other classifiers and accurately predicts the type of dementia in 68.22% of cases. Furthermore, the F1 score with respect to the control class was 83.72%, 64.8% with respect to the PCA class, and 47.69% with respect to the tAD class. The Cochran's Q test confirmed that there were differences between the performance of the models, $\chi2\ (4, N = 516) = 47.56$, $p < .001$.

Furthermore, analysing the confusion matrix of the winning model (the MLP classifier) in the ternary case suggests that the model misclassifies more often between the two types of dementia than with controls (see Table 4, Figure 5). This could be because people with dementia share some similar symptoms no matter the type and therefore their gait is much more similar to each other than to that of controls. Moreover, it is more common for the model to confuse participants with tAD with the control group than it is for the model to confuse participants with PCA with the control group. This could be because PCA affects visual processing more than tAD, and therefore the effects of this disease are more prominent in a trial such as this. This trend has also been identified by previous research done in the same programme of work at Pedestrian Accessibility Movement Environment Laboratory (PAMELA), which found that participants with early stage PCA performed worse than people with tAD (Yong et al. 2020). Therefore, because the gait of participants with PCA is more easily distinguishable from 'normal' gait than the gait of participants with tAD, the model does not confuse PCA with controls as often as it confuses tAD with controls.

In summary, these models could enable an in-the-wild screening tool for dementia, allowing people to conduct an initial screening, with reasonably high accuracy, before potentially receiving a clinical test to verify this. However, further research is required before this is possible, particularly in the case of the type-based classification where accuracy for the two types of dementia is lower than that for controls, suggesting that the current system may be sensitive to dementia, but not its type. See the discussion for more details.

Direction of Traversal and System Analysis

Analysis of number of sensors

A Kruskal-Wallis H test showed that there was a statistically significant difference in the importance of the sensors, $\chi2(6) = 157.13$, $p < .001$. Specifically, we tested the performance across model variants that used all different combinations of sensors (left foot; right foot; pelvis; left foot and right foot; left foot and pelvis; right foot and pelvis; left foot, right foot and pelvis). Post-hoc analysis showed the best performing combination was found to be the left and right foot sensor features together. These together gave a mean rank of 163.22 and an average accuracy of 85.94%. In contrast the worst performance was given by the pelvis features alone which had a mean rank of 13.00 and an accuracy of 74.45%. The importance of the placement and number of sensors, as given by the resulting accuracy, are given in Table 5.

The importance of the feet sensors in predictions could be explained simply because gait, which is heavily based on steps, can be more easily deduced from the movement of the feet, than the pelvis. Therefore, the accuracy of the model that uses a sensor on each foot is significantly higher than the others. Furthermore, it is interesting to note that the model that uses all three sensors yields a significantly lower accuracy than the model that uses only just two sensors – one on each foot. A potential reason for this is that given the data from each foot sensor, the pelvis sensor provides little additional useful information. Therefore, this information does not enhance the performance of the model, but could allow the model to identify trends that exist in the training set (or a subset of it) but do not generalise to other cases, causing the model to overfit to the training data.

The rest of the analyses (presented in this paper) used only the sensors attached to the feet as these produced the best performance. This analysis shows that when the data from sensors is processed independently of each other, sensors attached to participants' feet are more informative for making predictions.

These results of this analysis could not only be interesting to clinicians, and other researchers aiming to build similar systems, but also means that the sensor system can be truly unobtrusive as it does not require a pelvis sensor that can cause discomfort, thereby allowing its use in the wild. See the Discussion for more information about this.

The importance of features

Further analysis of the models was conducted to better understand how features from the gyroscope and the accelerometer contributed to the overall prediction (Figure 6). This was analysed by looking at the feature importance, using the tree-based models. Feature importance was calculated as the reduction in impurity (or error) that each node (or partition) provides weighted by the probability of reaching that node in the tree and then averaged over all trees. A Kruskal-Wallis H test showed that linear acceleration was statistically more importance than angular acceleration $\chi 2(31) = 795.47$, $p < .001$. While there is no conclusive explanation for this it is possible that this occurs because acceleration and velocity are directly related. Therefore, acceleration provides the model with useful information about the speed of a participant, the points when the foot is at rest, and how quickly the participant progresses through the trial. These have been identified by previous research (Verghese et al. 2007; del Campo et al. 2016; Carton et al. 2016; Castrillo et al. 2016; Montero-Odasso 2016; Cedervall, Halvorsen, and Åberg 2014) as factors that help distinguish participants with dementia from those without.

Furthermore, it appears (Figure 6, Table 6) that if we divide the trial into two halves (windows 1-4 and 4-8 respectively), then the second half appears more important generally for the model. To analyse this further the importance of the linear accelerations and the angular accelerations for the 4 windows in the two halves were summed together for each sensor and each type of acceleration. A second Kruskal-Wallis H test was applied followed by pairwise post-hoc Dunn tests with Bonferroni adjustments. Each of the pairwise comparisons was significant. The importance of the linear acceleration in the second half of the trial was found to be significantly greater than that of the first ($p=0.014$), which in turn was found to be significantly greater than the angular acceleration in the last half ($p<.001$). The angular acceleration in the first half was the least important and significantly less than the angular acceleration in the second half ($p=0.014$).

This analysis was conducted on all tree-based models (in both the binary and multi-class settings) which provide easy ways to calculate and analyse the importance of features, as well as being among the best performing models, and the trends identified across all these tree-based models were similar.

Therefore, this analysis identified the most informative components of the trial for distinguishing participants with dementia from controls, however, further research is required to provide an explanation for why these trends occur.

The effect of traversal direction

The analysis of the direction of traversal of the stairs that helps distinguish people with dementia from controls is presented in this section. The mean accuracy of the upward or downward directions are given in Table 7. This suggested that for people with dementia the binary models were more accurate in the upwards direction as compared to the downwards direction.

To analyse this further, the same analysis was conducted in the multiclass setting with accuracies split according to the class. The results of this analysis are summarised in Table 8.

A Friedman's Two-Way Analysis of Variance was conducted which proved there was a significant difference between the models and between up and down conditions $\chi 2(17) = 415.41$, $p < .001$. Pairwise analysis across two independent variables (models and up/down) was not conducted as it was thought to be over analysis of the data. However, from Table 8 it can be seen that in the multiclass tree-based models the percentage of the trials that were correctly classified as PCA is generally higher in the downward direction, which is in contrast to the results found for classifying dementia with binary models. This could be attributed to the fact that on the way down, the stairs are not directly in participants' line of sight when looking forward and, therefore, it is harder for them to process this information. Alternatively, it could be that descending stairs is less physically demanding, but the consequence of falling is greater when descending, causing anxiety in the participants.

While this analysis provides interesting insights into which direction of traversal is more informative for predicting dementia, the varied results across different models led to this analysis being inconclusive. Moreover, further research is required to provide an explanation for these differences.

The analysis of the importance of features and the direction of traversal provides some initial insights into how the gait of people with dementia (both PCA and tAD) could differ from that of controls, which may be informative to healthcare workers and patients. However, further analysis is required into the varied results and generalisability of these findings to other environments. See the Discussion for more details.

Discussion

This section discusses the contributions made, current limitations and future possible use cases of the STEP-UP system.

Detection and Discrimination of Dementia

While previous research has helped to identify the changes in gait that occur during the decline into dementia, the research has ignored two important factors that would allow such analyses to be used in the real world. Firstly, previous research relies heavily on experiments conducted in laboratory settings, using technologies such as optical systems that cannot be used in the real-world (Callisaya et al. 2017; Verghese et al. 2007; Wittwer, Webster, and Hill 2013) and treadmills which constrain the way of walking to a straight line. This limits the applications of this research as people hoping to use this method to screen for early cues of dementia would need to be subjected to these laboratory tests. Secondly, previous research often ignores different types of Alzheimer's focusing instead on tAD.

The use of low-cost wearable technology offers the opportunity to gather data about people's ability to conduct everyday tasks, including climbing or descending stairs as they go about their life. Previous research (Plant and Barton 2020) suggests that data from everyday life are more informative about a person's disease than data in clinical assessment laboratory where people may attempt to over control their behaviour. In addition, as such sensors get integrated into people's clothes and accessories, early detection of possible problems (especially rarer types of dementia like PCA) could be detected before people purposely look for a dementia assessment.

Our study has demonstrated the feasibility of deploying low-cost sensors to measure gait patterns for predicting dementia (both tAD and a rarer type of dementia: PCA) in everyday tasks of climbing and descending stairs. We have achieved this by focusing on low-level input features and investigating their non-linear mapping onto types of dementia and controlled groups with supervised classifiers. This is of critical importance when it comes to low-cost systems being used in the real world as calculating hand-engineered high-level gait features (e.g. (Verghese et al. 2007)) is often infeasible and requires high level controls. Also, low-level features used with artificial neural networks have been shown repeatedly to have higher robustness for other sensing modalities (Cho, Julier, and Bianchi-Berthouze 2019; Kostek, Szczuko, and Zwan 2004).

In this research we analysed the detection of dementia as compared to healthy participants, however, real-world deployment could enable larger datasets. This could further lead to an improvement in the performance not only on the detection of dementia cues but also on discriminating between different types of dementia. Moreover, the inclusion of more varied data such as that of participants with Mild Cognitive Impairement or early stages of dementia could enable this system to be used by these populations, allowing for early-stage detection. While we did not look at these populations, previous research analysing gait using similar methods and measures has found that gait is sensitive to early signs of dementia and can predict cognitive decline (Verghese et al. 2007; Cedervall, Halvorsen, and Åberg 2014; L. Wang et al. 2006; Waite et al. 2005; Marquis et al. 2002; Schaat et al. 2020; Halloway et al. 2019; Gwak, Woo, and Sarrafzadeh 2018). Therefore, deployment of this system in real-world settings could enable dementia detection in everyday settings which could bring several use cases and potential benefits. While in-depth analysis of this is left to future research, some of the potential future examples are discussed below:

Screening Tool for healthcare workers and practitioners: A screening tool which could be deployed in clinical settings or as an at-home test can be developed. The clinical tool could be used by community healthcare workers as well as general practitioners to enable easy detection of typical and atypical presentations of Alzheimer's disease. Carers' wellbeing can often be neglected, however they are often under considerable stress (Gilhooly et al. 2016). The amount of stress carers experience decreases with acceptance of the diagnosis and social support networks, and is increased with wishful thinking, denial and avoidance strategies (Gilhooly et al. 2016). An early diagnosis gives more time for acceptance and support networks to be established. These benefit the person diagnosed, their families and carers. It could be that beyond the benefits of simple screening we could also investigate ways of developing support tools for the carers, which could be linked to the stage of dementia of the person for whom they caring.

General self-screening: As sensors are increasingly integrated into our daily activities (e.g., sensor in shoes for running, imaging for fitness tracking) and used to quantify our wellbeing (Cho 2021; Cho et al. 2017), such sensors could be used together to detect and identify cues of decline and dementia. Our results provide some insights on how the sensors could be used in the wild. Firstly, our research found that the presence of dementia is more easily detected during upwards stair climbing,

suggesting that the gait of people with dementia is more abnormal during upwards stair climbing. The same sensors placed on the shoes could first detect upward stair climbing (Formento et al. 2014) and data from this activity can be prioritised for more accurate predictions. Similarly, the sensors could also detect long periods of activity and even fatigue or pain (C. Wang et al. 2019) and consider such variables when evaluating the assessment tool outcome. Finally, as any motor activity modelling suffers from people's idiosyncrasy, such models could take advantage of the long history of sensor data gathered from the person to build personal models of what is a normal pattern (given the physical ability including vision of the person) and hence detect possible sudden declines that may indicate such underlying causes of dementia and even atypical causes.

Support Tool for patients: It would seem feasible to also develop the ability to classify deteriorations in a person's condition following diagnosis. This would need a larger data set collected in the wild. Once developed decline in gait such as those detected by lab-based studies (e.g. (Callisaya et al. 2017; Verghese et al. 2007)) could be detected as people conduct their daily activities and be directly linked to clinical care pathways. This would enable person-centred care to be established, rather than simply asking people to return for appointments based on standard time predictions of decline.

An important perspective is on the effect of different combinations of sensors on the detection performance. Our research found that of all combinations of the sensors, models using only the sensors attached to the feet performed best. This led to us dropping the pelvis sensor from further analyses. Additionally, a sensor constantly attached to a person's pelvis may cause discomfort. Therefore, our research suggests that a truly unobtrusive system could be built simply with sensors attached to people's shoes. Furthermore, the support tool could be further developed to be predictive of decline, providing further support to people with dementia and their care givers.

Limitations

Despite promising results, there is room for improvement. We discuss points to help the deployment of such a system.

Discriminating Type

While the model has shown a good performance (from LOSO cross-validation) in the multi-class classification (Control vs tAD vs PCA), we have found lower performance in discriminating the two types of dementia when samples from the controlled group are not considered in the classification task. This can provide insights. First, this could be related to the fact that the gait of the two subtypes of dementia was very similar to each other, suggesting that gait is sensitive to dementia as a whole, but less sensitive to the type of dementia. This could suggest that different measures may be required to provide a more comprehensive diagnosis. For example, in PCA vision is predominantly affected with memory often being (initially) unaffected. Second, the data from healthy participants could play an essential role in discriminating patterns associated with each dementia type. Third, when it comes to the dementia detection task (dementia vs. control), the proposed system results in a very high accuracy of 87.02%.

Generalization issues and Dataset

Another potential limitation in this study is that models might be overfit to the data, reducing its ability to generalise to unseen data. While we prevented this as much as possible by using LOSO validation, ensuring the model was not only tested on unseen data but on data from an unseen participant. However, all the data from all participants was collected on the same staircase using the

same system setup to collect the data. Therefore, these models may not generalise to other environments, other staircases or other IMU systems. This may limit the direct application of this system to the real-world diagnosis of dementia. Therefore, further research is required to prove the generalisability of this research to other environments and system implementations.

Another related issue was that it was more difficult to achieve a high degree of control in the task especially in people with dementia. This may have resulted in patients taking breaks in the middle of the task, not initially standing in the correct start position, etc. Therefore, the model might use these artefacts to discriminate patients from controls rather than their gait. Nevertheless, these behaviours are symptoms of dementia that should generalise across patients.

Furthermore, in this study we only compared the gait of participants with dementia to healthy age-matched controls. Therefore, this model may be overfit to distinguishing healthy and unhealthy participants and may not be able to distinguish dementia from other diseases with similar presentations or people with a bad physical condition. Therefore, this requires further research and fine-tuning of this issue. We believe that the deployment of this system in the real-world would enable overcoming these overfitting issues by allowing more varied data to be tested.

Conclusion

This research demonstrates the feasibility of automatically detecting both the more typical Alzheimer's Disease (tAD) as well as a rarer and distinct form of dementia – Posterior Cortical Atrophy (PCA) – based on gait in a real world-environment. To this end, we propose the use of low-level features based on windowed averaging of data from a low-cost, unobtrusive IMU system. These features are easy to calculate from a small number of IMU sensors, enabling their use in a real-world system. We also demonstrate that these features can be used with Machine Learning models to predict dementia with 87.02% accuracy. Furthermore, we demonstrate that a sensor placed on each foot is sufficient for this analysis. Lastly, we demonstrate the models are better able to discriminate people with dementia from healthy controls when they are climbing up stairs, suggesting that people with dementia find it harder to climb up stairs.

Therefore, this research concludes that machine learning analysis of IMU data, gathered from a person's gait in a real-world environment, could unobtrusively be used to assess the risk of having dementia. Once further researched, a system such as this could provide an initial assessment of the risk of having a certain type of dementia before conducting any clinical tests, thereby streamlining and enhancing the diagnostic process. Therefore, not only are these results interesting from a research perspective, but also have potential real-world applications.

Figure Labels

Figure 1. Project STEP-UP: to enable low-cost and wearable IMU sensors to infer dementia types in the wild whilst climbing stairs.

Figure 2. Technical details of Step-up framework: (A) Gait Recording procedure using wearable IMUs, (B) Procedure for the exclusion of corrupted files, (C) Feature extraction procedure using windowed averaging, (D) Model training and tuning procedure, (E) Validation procedure using Leave One Out Validation. The procedure for splitting the dataset into training and testing sets is shown under (D) and (E).

Figure 3. A plot of the prediction accuracies of the Random Forest Classifier when using different numbers of windows (1 to 15) for constructing the features.

Figure 4. A confusion matrix for the binary Random Forest model

Figure 5. A confusion matrix for the ternary MLP model

Figure 6. The Importance of the Features for the Random Forest Classifier when predicting Dementia. The features used were the windowed averages (number of windows 8) of linear acceleration (blue bars) and angular acceleration data (orange bars) for both the left (left hand side) and right sensor (right hand side). Feature importance was calculated as the reduction in impurity (or error) that each node (or partition) provides weighted by the probability of reaching that node in the tree and then averaged over all trees.

Tables

Table 1. The dataset before the removing the corrupted files compared to the dataset after this removal.

Group	Number of trials (before removal)	Number of trials (after removal)
Control	208	207
PCA	159	150
tAD	160	159
Total	527	516

Table 2. Values of the hyper-parameters (for each model) that yielded the highest performance and were used in all analyses.

Model	**Parameter Name**	**Binary Parameters**	**Multiclass Parameters**
	Number of trees	80	70
	Maximum depth of trees	1	3
	Minimum samples in leaf nodes	2	2
Gradient Boosting	Learning rate	0.15	0.05
Random Forest	Number of trees	120	120

	Maximum depth of trees	None	3
	Minimum samples in leaf nodes	5	2
MLP	Number of units in hidden layer	8	8
		Non-linearity	Logistic/Sigmoid Function
	Maximum number of iterations	750	750
	Learning rate	0.0002	0.0002

Table 3. Results from a representative run of the models for detecting the dementia (PCA/tAD).

Model	Accuracy	F1 Score (wrt the Control class)	F1 Score (wrt the Dementia (PCA/tAD class)
Gradient Boosting	86.05%	82.78%	88.27%
Random Forest	87.02%	83.14%	88.38%
MLP	86.63%	82.71%	87.75%

Table 4. Results from a representative run of the models for detecting the type of dementia.

Model	Accuracy	F1 Score (wrt the Control class)	F1 Score (wrt the PCA class)	F1 Score (wrt the tAD class)
MLP	68.22%	83.72%	64.8%	47.69%

Table 5. Average accuracies of Binary Gradient Boosting Classifiers using different sensors. The table shows the average accuracies (across 25 samples) of the Binary Gradient Boosting classifier when using the data from different combinations of the sensors to construct the features.

Position of Sensors Used	Accuracy
Left Foot	81.99%
Right Foot	84.78%
Pelvis	74.45%
Left, Right Foot	85.94%

Left Foot, Pelvis	81.90%
Right Foot, Pelvis	83.25%
Left Foot, Right Foot, Pelvis	83.41%

Table 6. Results of hypothesis testing comparing the linear and angular acceleration in the first (windows 1-4) and second (windows 5-8) halves of the trial.

	First Half	Second Half	p-value
Linear Acceleration	24.17%	42.25%	<0.001
Angular Acceleration	19.20%	14.37%	<0.001

Table 7. Results of hypothesis testing comparing the prediction accuracies attained in the upward and downward directions.

Model	Upward Accuracy	Downward Accuracy
Random Forest	86.77%	85.31%
Gradient Boosting	86.97%	86.08%
MLP	89.29%	82.79%

Table 8. The average accuracies (across 25 samples) of the better performing models for predicting dementia phenotype.

	Upwards Accuracy			Downwards Accuracy		
Model	**Control %**	**PCA %**	**tAD %**	**Control %**	**PCA %**	**tAD %**
Random Forest	80.12	56.43	51.1	79.46	63.11	41.16
Gradient Boosting	79.03	61.95	45.90	89.69	71.37	30.03
MLP	79.42	74.32	49.5	92.19	71.9	34.89

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

CH: conception of analysis, acquisition of data, drafting manuscript, analysis of data, built use cases; WB: conception of analysis, analysis of data, drafting manuscript, built use cases; KY: conception and design of experimental protocol, acquisition of data, assisted drafting manuscript; IM, TS: acquisition of data, advised on data analysis; AC, BY: acquisition of data; RS, DB: conception and design of experimental protocol; NT: conception and design of experimental protocol; SC: conception and design of experimental protocol, assisted drafting manuscript; NB: drafting manuscript, advised on analysis of data, built use cases; YC: overall technical supervision, drafting manuscript, advised on analysis of data, built use cases.

Funding

The Dementia Research Centre is an Alzheimer's Research UK Co-ordinating Centre and is supported by Alzheimer's Research UK, Brain Research Trust, and The Wolfson Foundation. This work was also supported by the NIHR Queen Square Dementia Biomedical Research Unit and by an Alzheimer's Research UK Senior Research Fellowship (ART-SRF2010-3) and ESRC/NIHR (ES/L001810/1) and EPSRC (EP/M006093/1) grants to SC.

Acknowledgements

We would like to thank participants for their patience and goodwill in taking part. The Dementia Research Centre is an Alzheimer's Research UK Co-ordinating Centre and is supported by Alzheimer's Research UK, Brain Research Trust, and The Wolfson Foundation. This work was also supported by the NIHR Queen Square Dementia Biomedical Research Unit and by an Alzheimer's Research UK Senior Research Fellowship (ART-SRF2010-3) and ESRC/NIHR (ES/L001810/1) and EPSRC (EP/M006093/1) grants to SC.

Data Availability

The datasets ANALYZED for this study can be found in the *Seeing what they see: Compensating for cortical visual dysfunction in Alzheimer's disease 2014-2018* http://reshare.ukdataservice.ac.uk/853147/.

References