Cognitive Computation (2024) 16:1004–1021
https://doi.org/10.1007/s12559-024-10279-z

Cognitive Impairment Detection Based on Frontal Camera Scene While Performing Handwriting Tasks

Federico Candela[1] · Santina Romeo[1] · Marcos Faundez-Zanuy[2] · Pau Ferrer-Ramos[2]

Received: 5 September 2023 / Accepted: 26 March 2024 / Published online: 10 April 2024

Abstract
Diagnosing cognitive impairment is an ongoing field of research especially in the elderly. Assessing the health status of the elderly can be a complex process that requires both subjective and objective measures. Subjective measures, such as self-reported responses to questions, can provide valuable information about a person's experiences, feelings, and beliefs. However, from a scientific point of view, objective measures, based on quantifiable data that can be used to assess a person's physical and cognitive functioning, are more appropriate and rigorous. The proposed system is based on the use of non-invasive instrumentation, which includes video images acquired with a frontal camera while the user performs different handwriting tasks on a Wacom tablet. We have acquired a new multimodal database of 191 elder subjects, which has been classified by human experts into healthy and cognitive impairment users by means of the standard pentagon copying test. The automatic classification was carried out using a video segmentation algorithm through the technique of shot boundary detection, in conjunction with a Transformer neural network. We obtain a multiclass classification accuracy of 77% and two-class accuracy of 83% based on frontal camera images, which basically detects head movements during handwriting tasks. Our automatic system can replicate human classification of handwritten pentagon copying test, opening a new method for cognitive impairment detection based on head movements. We also demonstrate the possibility to identifying the handwritten task performed by the user, based on frontal camera images and a Transformer neural network.

Keywords Deep learning · Neurodegenerative disorders · Mild cognitive impairment · Eye movement · Image processing

Introduction

Assessing the health status of the elderly can be a complex process that requires both subjective and objective measures. Subjective measures, such as self-reported responses to questions, can provide valuable information about a person's experiences, feelings, and beliefs. However, from a scientific perspective, objective measures, based on quantifiable data that can be used to assess a person's physical and cognitive functioning, are more appropriate and rigorous. The use of multiple measures can help provide a more accurate and reliable assessment and can improve the accuracy of diagnoses and treatment plans. In this paper, we turn our attention to objective measures.

Examples of objective measures are physical tests, laboratory tests, and cognitive assessments. To assess the health condition of the elderly, neurodegenerative diseases may vary depending on the specific condition and assessment goals. In recent years, several studies have been conducted to explore the use of artificial intelligence (AI) in objective measures for the diagnosis and assessment of neurodegenerative diseases.

Current research shows that machine learning algorithms have been developed for analyzing brain images (such as MRI or PET) to identify patterns or hallmarks of neurodegenerative diseases. These algorithms can help detect structural abnormalities, such as brain atrophy, or identify

✉ Marcos Faundez-Zanuy
faundez@tecnocampus.cat

Federico Candela
federico.candela@unirc.it

Santina Romeo
santina.romeo@unirc.it

Pau Ferrer-Ramos
pferrerr@tecnocampus.cat

[1] University Mediterranea of Reggio Calabria, Via dell'Universit`a, 25, City 8914, Italy

[2] Tecnocampus, Universitat Pompeu Fabra, Avda. Ernest Lluch 32, Mataro 08302, Spain

characteristic features of specific conditions such as Alzheimer's disease (AD) [1]. AI has also been implemented to analyze EEG data to identify patterns or signatures of brain activity associated with neurodegenerative diseases. In Doulamis and Voulodimos [2], the authors used deep learning to classify EEG signals from patients with mild cognitive impairment (MCI) and AD. Machine learning algorithms were used for feature extraction and classification of EEG data. Distinctive brainwave traits can be correlated with specific patient conditions.

Both early detection of cognitive impairment and cognitive impairment assessment pose unique challenges, but early detection is often considered more difficult and more important [3] for several reasons:

- Timely intervention and treatment: Detecting cognitive impairment in its early stages allows for prompt intervention and the implementation of appropriate treatments [4]. Certain cognitive disorders, such as Alzheimer's disease, may benefit from early pharmacological or non-pharmacological interventions, potentially slowing down the progression of symptoms.
- Improved quality of life: Early detection enables individuals to receive timely support and resources to cope with cognitive changes. This can enhance their overall quality of life by providing them with tools and strategies to manage cognitive challenges and maintain independence for as long as possible.
- Reduced caregiver burden: Identifying cognitive impairment early allows caregivers to plan and adapt to the evolving needs of the individual. This proactive approach can reduce caregiver burden by facilitating better preparation, support, and the development of coping mechanisms.
- Patient and family empowerment: Early detection empowers individuals and their families with knowledge about the condition. It allows for informed decision-making regarding future care plans, legal matters, and financial arrangements, promoting a sense of control and autonomy.
- Facilitation of research: Early detection contributes valuable data for research purposes, aiding scientists and healthcare professionals in understanding the progression of cognitive disorders. This, in turn, can lead to the development of more effective treatments and interventions.

Early detection of cognitive impairment is challenging due to several facts, such as it has to cope with subtle symptoms. Early-stage cognitive impairment often presents with subtle and non-specific symptoms, making it challenging to differentiate from normal age-related cognitive changes or other health issues. On the other hand, the individuals in the early stages of cognitive impairment may lack awareness of their condition, hindering self-reporting and self-recognition. This reliance on external observation adds complexity to the detection process. And last but not least, diagnosing cognitive impairment at an early stage involves a degree of uncertainty due to the potential for variations in symptoms and the absence of clear-cut diagnostic criteria. This complexity can pose challenges for healthcare professionals.

In this paper, we explore the use of a new signal to detect the presence of cognitive impairment in a user. It is a non-invasive methodology as the acquired signals are eye tracking and head movements detected by a frontal camera available in wearable commercial eye-tracker system. These signals are acquired while performing several writing tests on a WACOM Cintiq tablet. All of the user's eye movements are recorded using the commercial eye tracking device "Tobii Pro Glasses 3 ®" Tob [5], shown in Fig. 1. This device is worn like a normal pair of glasses and is operated by a battery without any connection to a computer. The glasses record a detailed first-person perspective (from frontal camera) as well as participant's gaze in real time without intruding on their natural behavior.

The hypothesis of the paper is that the images acquired from a frontal camera while performing handwritten tasks can reveal the presence or absence of cognitive impairment. This will open the possibility for a better multimodal detection based on video signals in addition to the well-known handwritten signals [6]. For that purpose, we have updated the handAQUS acquisition software Mucha [7] to acquire simultaneously both signals from the same software. This permits the time synchronization of both signals.

It is noteworthy that Ray-Ban® and Meta® have recently introduced smart glasses capable of capturing audio and video. These glasses enable users to directly live stream what they see and hear on platforms like Instagram and Facebook. Consequently, the emergence of innovative technological applications based on these signals is likely to accelerate in the near future, facilitated by the availability of new hardware for domestic users.

Fig. 1 Tobii Pro Glasses 3 ®

State of the Art

Cognitive impairment refers to problems with learning and memory, language, executive function (managing daily work and life), attention, perceptual motor skills (interacting with the environment), and social cognition (interacting with other people). There is a wide spectrum of cognitive impairment in adults that ranges from mild (barely noticeable) to full-blown dementia (most commonly Alzheimer disease). With milder forms of cognitive impairment, changes are often not noticed by patients or their friends or family. Therefore, screening tests for cognitive impairment are of interest to primary care clinicians, especially if they are fast, cheap, and non-invasive.

Many brief screening tests for cognitive impairment are available. Commonly used tests include the Mini-Mental State Examination (MMSE) [8]. The test includes questions in a number of areas including attention, calculation, comprehension, construction, naming, orientation, recall, registration, repetition, spelling, and writing. After the MMSE, clock drawing test (CDT) is the second most widely used test for grading cognitive states. The origin of the CDT is not clear. Evidence suggests that it was first used by the British neurologist/psychiatrist Sir Henry Head [9]. In the original testing method, a pre-drawn clock is given to the subject, who is asked to draw the clock hands indicating 10 min past 11 o'clock. Based on our experience, the easier to instruct the task the better. Some tasks are time consuming because:

(a) The users do not perfectly understand what they must do. For instance, how much time must I stay producing consecutive circles? Should I produce my lines over the Archimedes spiral template or between lines? Which is the starting point? Etc. Some examples of different handwritten tasks are available in [10]
(b) The users have not been exposed or they are not habituated "to see" the object they have to write. One example is the analogue CDT in young people, as they check the time in digital clocks, smartphones, etc.

Probably for these reasons, the pentagon copying test is one of the most popular. They do not need to produce a drawing of something they should remember. They just need to copy a model, and they can start from the place they want.

In general, screening tests generally involve asking patients to perform a series of tasks that assess one or more aspects of cognitive function. A positive screening test result leads to additional testing for dementia that can include blood tests, magnetic resonance imaging of the brain, and more in-depth neuropsychologic testing by specialists.

The population under consideration for screening for cognitive impairment are mainly adults aged 65 years or older who live in the community (i.e., not in a nursing home) and do not have any signs or symptoms of cognitive impairment.

To the best of our knowledge, there is no scientific paper devoted to head movement acquisition while performing a handwriting task on a digitizing tablet. However, some papers exist based on head movements and cognitive impairment. Head turning is an easily observed and categorized sign and may raise suspicion of the presence of a cognitive disorder Ghadiri-Sani et al. [11]. In Durães et al. [12], the authors discuss that the head turning sign (HTS) is frequently noticed in clinical practice. In their paper, a total of 84 patients were analyzed. They found that HTS was more prevalent in AD than in MCI or in FrontoTemporal Dementia (FTD). It also correlated negatively with cognitive measures and depression. They conclude that the presence of the HTS in a cognitively impaired individual suggests a diagnosis of AD. A higher HTS frequency correlates with higher cerebrospinal fluid (CSF) total tau levels, a smaller Geriatric Depression Scale (GDS) score Yesagave et al. [13], and worse cognitive measures. In the MCI subgroup, the HTS may suggest a higher risk of progression.

Eye movement (EM) is also related to cognitive impairment, according to the scientific literature. Opwonya et al. [14] conclude that EM metrics combined with demographics and cognitive test scores enhance MCI prediction, making it a non-invasive, cost-effective method to identify early stages of cognitive decline. They computed EM metrics from participants who completed the ProSaccade (or Go condition)/AntiSaccade (PS/AS) and No-go tasks (gaze focused on the center of a screen ignoring objects on the left and right side).

Cognitive and Biologically Inspired Approach

The proposed system is rooted in a cognitive and biologically inspired approach to diagnosing neurodegenerative disorders, particularly in the elderly (all the samples were acquired from people over 60, with an average of 71 years old). The emphasis is on utilizing objective measures that involve quantifiable data related to physical and cognitive functioning. This aligns with the understanding that cognitive decline and neurodegenerative disorders often manifest in measurable changes in behavior, motor skills, and other physiological parameters.

The emphasis on head movements during handwriting tasks suggests an exploration of motor control and coordination, which are closely tied to cognitive processes. In fact, cognitive impairment often results in observable changes in fine motor skills, making this a biologically relevant aspect of assessment. On the other hand, the new multimodal database PECT-Tecnocampus especially acquired for this research proposal implies an integration of various data types, mirroring the complexity of cognitive processes in real-world scenarios. This approach draws inspiration from

the multifaceted nature of human cognition, as information from different modalities is often processed and integrated in the brain.

Automatic classification is proposed, based on deep learning. This is aligned with the biological inspiration drawn from neural networks in the human brain. The ability of deep learning algorithms to discern complex patterns and relationships is analogous to the neural processing capabilities observed in biological systems. Experimental results show that cognitive impairment may result in distinctive patterns that can be learned and recognized by machine learning algorithms. This approach is aligned with very recent publications, such as Howard [15].

Worth to mention that automatic detection of cognitive impairments enables ambient intelligence systems to offer personalized assistance and adaptation [16]: The system can adjust lighting, temperature, and other environmental factors to promote comfort and reduce confusion. Additionally, it can provide reminders for daily activities, medication, and appointments, catering to the specific needs of each individual.

This paper is structured as follows: The "Methodology" section introduces the methodology of the proposed system, the "Experimental Verification" section shows the experimental results obtained and related discussions, and finally, we conclude with the "Discussion and Conclusions" section.

Methodology

Head movements can provide important information about a person's motor control, balance, and coordination. There are studies suggesting that certain head movements, such as voluntary jerks [17] or tremors [18], may be indicative of neurological or movement disorders. However, it is important to note that head movements alone may not be sufficient to accurately diagnose a medical condition. To accurately diagnose a pathological condition requires a comprehensive evaluation including a series of physical and neurological tests. Nevertheless, the goal of this research is not to provide a medical diagnose. It is to propose a simple method to rise a flag when there is a suspicious case of cognitive impairment.

Pentagon Copying Test

In the Pentagon Crossing Test of the Montreal Cognitive Assessment (MoCA) Nasreddine et al. [19] individuals are asked to draw a specific geometric figure, a pentagon, while following a set of instructions. The test assesses visuospatial abilities, executive function, and attention. The individual is given a piece of paper with a pre-drawn pentagon and is instructed to connect alternate corners of the pentagon with straight lines. The goal is to accurately complete the drawing according to the provided instructions.

The test is scored based on the correctness of the drawing and the adherence to the instructions. Errors such as incorrect line placement, extra lines, or failure to follow the instructions can indicate difficulties in visuospatial processing and executive function, which are cognitive domains often affected in certain neurological conditions. Most of the experts assign a zero or one score depending on correctness. It is considered correct if there are two pentagons, each pentagon has five sides, and the intersection between both pentagons is correct. Otherwise, the score is zero. This is a fast an easy-to-follow process.

In some scenarios, a more quantitative evaluation is done, especially when implemented in a computerized version. For instance, in Nagaratman et al. [20], it is possible to assign a score of 1–10 for each portrayal. Table 1 summarizes the scoring system. The rotation of the figures or tremor was overlooked according to the original criteria.

Figure 2 shows an example of PDT for a user affected by AD in basal situation, 6, 12, and 18 months after diagnose.

In this paper, we have used the binary classification as our aim is cognitive impairment detection rather than assessment. We have manually inspected the pentagons produced

Table 1 Scoring system in a 1 to 10 scale for PDT

Score	Condition
10: normal	All sides were equal, all the angles of the figures were present, and the two figures intersected
9	One or two sides are of different length
8	Same as score 9 but no intersection
7	Loss of one or more angles
6	One pentagon incomplete
5	Reduced number of sides
4	Loss of sides and angles
3	Grossly incomplete sides
2	Not interpretable
1	No reasonable attempt at drawing or the drawing was just a squiggle or scrawl

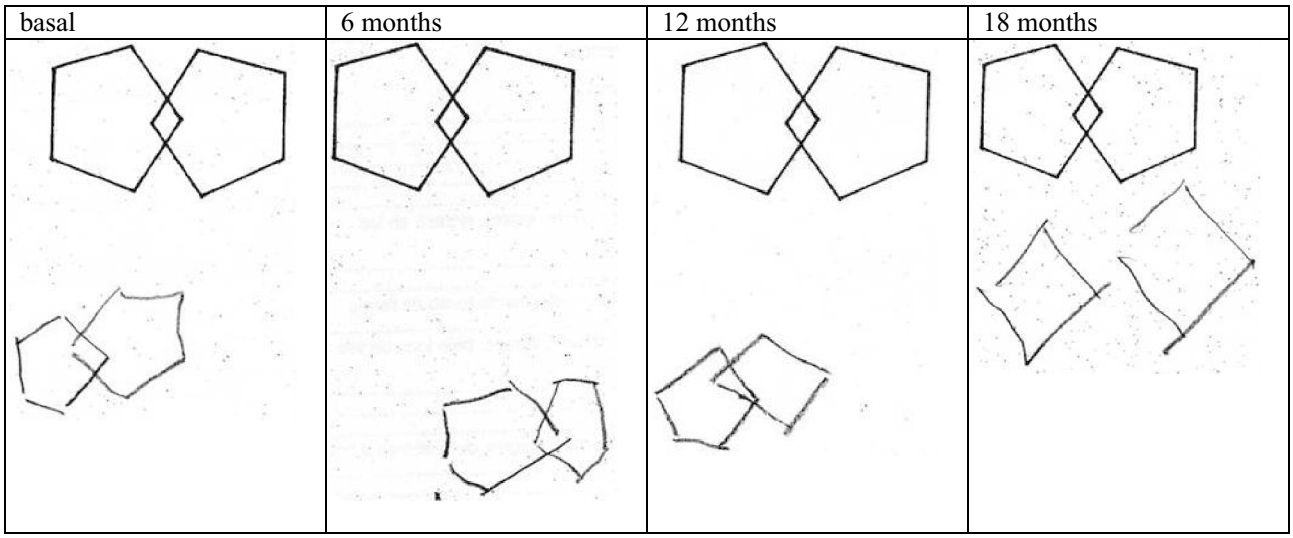

Fig. 2 PDT produced by a user affected by AD in basal situation, 6, 12, and 18 months after diagnosis

by all the users and assigned healthy control or cognitive impairment group.

Database

In this paper, we have acquired a new database called PECT-Tecnocampus. The acquisitions were carried out in several civic centers in the city of Mataró (Barcelona, Spain) and gathered 191 volunteers over the age of 60 from the Maresme region. All the donors signed an informed consent according to ethical regulations.

Individual physical abilities were assessed using questionnaires and physical tests (balance, mobility, cardiorespiratory fitness, among others), and an individualized report of the results was provided. Figure 3 shows the histogram of the age of donors split in males and females. Mean (m) and standard deviation (std) for ages of males and females are respectively: males ($m = 71.76$, $std = 5.64$), females ($m = 71.26$, $std = 6.28$).

From the 191 users, 174 of them finished all the handwritten tasks on a Wacom Cintiq digitizing tablet.

In our previous published databases, we acquired all the handwritten tasks in a single DIN A4 sheet. However, for this new database, we decided to use two different DIN A4 sheets: one for handwritten text and signature and another one for drawings. This permitted larger sizes and helped visual impaired people to finish the tasks. Thus, the handwritten tasks can be classified into two groups:

- Drawings: (a) two pentagon copy test, (b) house copy test, (c) spring drawing, (d) Archimedes spiral, (e) concentric circles performed at regular speed, (f) straight line connecting two dots without touching the lower and upper black bars. Figure 4 shows the template used for this tasks
- Handwriting: (1) signature performed two times, (2) words in capital letters copy, (3) cursive letter sentence copy. For more information on different handwritten tasks check [10].

PECT-Tecnocampus database was acquired with a Wacom Cintiq 16 tablet and a modified version of the original HandAQUS software Mucha [7]. The features of this tablet are 5080 lpi and 8192 pressure levels. From this database, a manual classification by visual inspection of the pentagon copying test (PDT) [8] was performed by a lecturer of

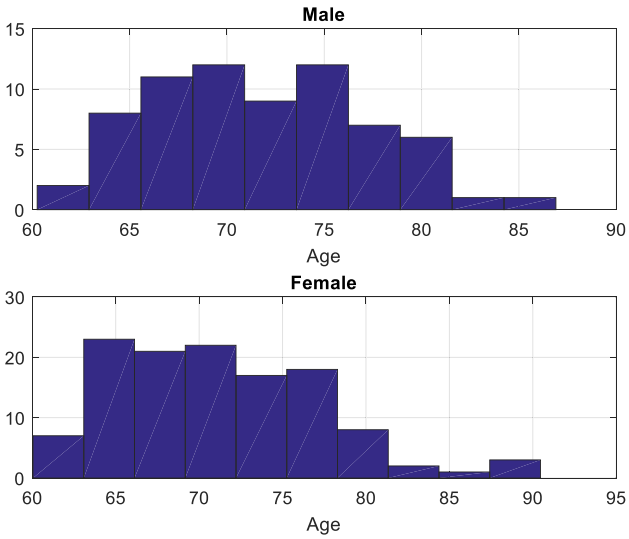

Fig. 3 Histogram of ages for males (top) and females (bottom)

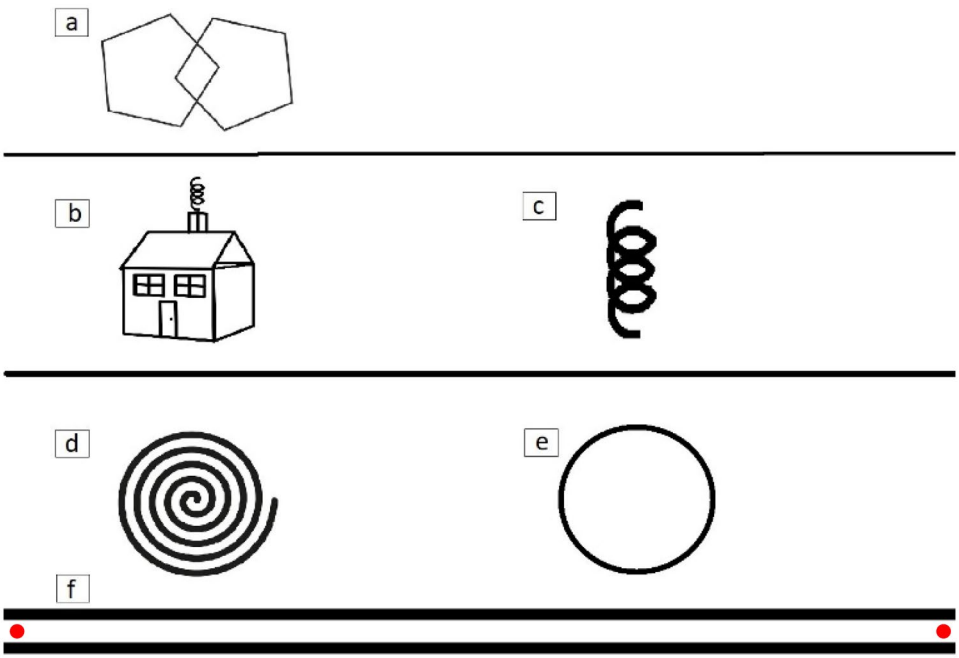

Fig. 4 Template used for drawing tests. It consists of six different tasks

the health sciences faculty at Tecnocampus. Each user was assigned to a cognitive impairment or healthy group. Those who fail to pass the PDT were classified as users with some cognitive impairment.

From the 174 users, 81 failed to pass the PDT (52 females and 29 males). The criterion to evaluate the quality of the PDT was the following:

- The PDT is successful if there are two pentagons that intersect at two points.
- Each pentagon must have exactly five sides and five angles and must interlock at two points of contact.
- It does not matter if the angles are not equal, although it is necessary that the pentagons are not open at any corner.
- Small errors are allowed when almost imperceptible, and also if tremors are evident, and the lines are not completely straight.

The MMSE alone does not provide any diagnosis, and although it is a useful tool when assessing to a patient with memory problems, the diagnosis of mild cognitive impairment (MCI) or dementia is done by complementing it with a good medical history in addition to a correct physical examination and performance of complementary tests. Even so, sometimes evolutionary monitoring of the patient is necessary to give a specific diagnosis.

Due to privacy issues and also due to the difficulty to know the pathologies of almost 200 users, each of them with a different family doctor, specialist, etc., we cannot know the exact pathology, if any, that affects each user. However, the goal of this study was not to diagnose patients. It was to study the health conditions of elder people including cognitive impairment.

The PDT is a sub-test of the Mini-Mental State Examination (MMSE) [8], used extensively in clinical and research settings as a measure of cognitive impairment. This manual classification is used as "ground truth" for automatic classification based on head movements acquired by frontal camera of eye-tracker system, described next.

While performing the handwriting tasks, the users wore the Eye-tracker Tobii Pro Glasses 3® (see Fig. 1). The glasses have several cameras, pointing to the user's eye (from glass to eye) and one camera pointing to the general scene seen by the user (from the glass to outside). Eye tracker provides a set of data that can be used to analyze visual behavior, reading habits, attention, interest, and other related metrics.

One key point was that in a database acquisition, one of the most difficult tasks is the recruitment of donors. However, once you get them, it is a good practice to include as much as non-invasive sensors as possible. For this reason, we added the eye tracker to the handwriting task acquisitions.

The initial idea was to use a wearable eye tracker while performing handwriting tasks and analyze the eye-tracking signals. This eye tracker not only acquires info related to gaze, pupil sizes, eye-closing, etc., but it also provides a video image recording of the scene seen by the user wearing the glasses. However, during the database acquisition, we detected calibration problems for users wearing corrective lenses. In this case, the users had to add the eye-tracker glasses over the corrective lenses used to overcome visual impairment. Unfortunately, the use of corrective lenses is

a quite usual situation for elder people when doing writing tasks.

Thus, the use of eye-tracker information from the current database would require to detect and probably remove from database and/or from the experiments those users with uncalibrated samples. However, we have not found failure to acquire issues with the frontal camera, which is also an interesting signal to analyze. Thus, we decided to propose a system based on this signal rather on eye-tracking ones, which are left for a future work (some acquired signals can exhibit low sensibility to calibration errors).

Worth to mention that wearing eye glasses such as Tobii Pro Glass 3 is quite comfortable for users and more convenient than some other specially designed and mounted devices on the head of the user.

For future work, we consider important to use the option of corrective lenses kit available from Tobii. The kit includes individual lenses for both left and right eyes ranging from − 8.0 to + 3 diopter in intervals of 0.5 diopters. This kit extends the applicability of Tobii Pro Glasses 3 to people with the most common forms of near- and farsightedness. However, this would increase the acquisition time per user as in addition to eye-tracker calibration, it requires another previous calibration for each user.

Frontal camera provides useful information about head movements during handwriting tasks. This camera provides 25 fps of 1920 × 1080 pixels each in RGB format.

Shot Boundary Detection with Background Subtraction

Shot boundary detection is a technique used to identify significant changes between shots in a video. This method relies on the pixel difference between two consecutive shots to determine if a scene change has occurred. Background subtraction is used to separate the foreground (moving objects) from the static background in an image or sequence of images. The algorithm can be found in Candela et al. [21]

Through this process, our method not only accurately identifies scene changes but also enables the extraction and saving of video segments corresponding to individual scenes, thereby providing a powerful tool for detailed analysis and segmentation of complex video content.

Video Dataset Creation with Shot Boundary Detection

To develop our training dataset, we employed the shot boundary detection technique to segment a continuous video capturing the entire user test into six separate subsequences. Each sub-sequence corresponds to a different task on the Cintiq tablet: drawing a pentagon, house, vertical spiral, Archimedean spiral, concentric circles, and a straight line. The ability of shot boundary detection to pinpoint scene changes tied to shifts in activity enabled us to isolate each graphic task into an individual video.

Video analysis revealed that users with cognitive deficits exhibited irregular eye and head movements, particularly an inability to maintain a steady gaze on the pen while performing the tasks.

The data subsequently were labeled by hand creating two macro-categories (pass/not pass pentagon test) including six tasks for each macro-category.

As a result, we categorized the training data based on the six tasks performed by the two types of users: six classes for those who passed the pentagon test and another six for those who failed, making a total of twelve categories.

Shot Boundary Transformer Detection

Transformers, proposed by Vaswani et al. [22], are particular deep learning models. The characteristic feature of these networks is self-attention, a process of differentially weighting the meaning of each part of the input data by working predominantly on sequential data.

For all sub-videos, a feature extractor is constructed using the DenseNet121 [23] model pre-trained on ImageNet [24], which is used to extract features from video frames. Each sub-video input, labeled as S_i, is transformed into a three dimensional matrix X_i, with dimensions $t \times h \times w$ (t = time, h = height, w = width). This conversion process is depicted in Fig. 5, where each array X_i is an element of the sub-video input S_i.

The input is parsed into i space-time tokens thanks to an encoder. The encoder consists of an encoding layer that processes the input iteratively one layer after another, allowing it to draw from the state at any previous point along the sequence. The encoder consists of two main components: a self-attention mechanism and a feed-forward neural network.

As the first step, for each level of attention, the transformer learns three weight matrices by defining: W_Q query weights, W_K the key weights, and W_v the value weights.

For each token, the input data embedding Xi is multiplied by each of the three weight matrices to produce the three corresponding vectors: $q_i = Xi \cdot W_Q$, query weights; $k_i = X_i \cdot W_K$, the key weights; $v_i = X_i \cdot W_v$, the value weights.

In the second step, the attention weights aij are calculated (Eq. 1) using the dot product between the query vector and the key vector for each token pair and are normalized by dividing by the square root of the dimension of the key vectors ($d_k = t \times h \times w$). This helps in stabilizing the gradients during training. The output for a token i is the weighted sum of the value vectors of all tokens, weighted by the normalized attention weights aij:

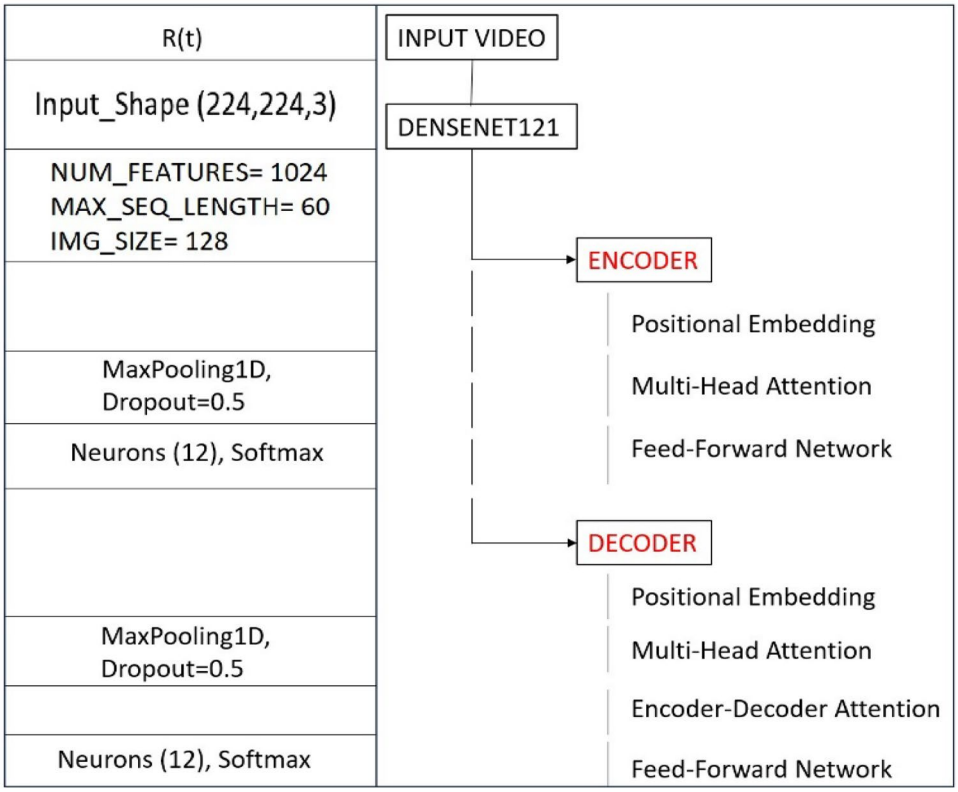

Fig. 5 Network architecture

$$a_{ij} = \frac{W_Q \cdot W_K}{\sqrt{d_k}} \tag{1}$$

In the third and final step, the encoder's output is used as input for a multi-head attention layer, which then feeds data to the decoder. Each "attention head" includes a set of matrices (W_Q, W_K, W_v) and allows for the computation of attention across different subspaces simultaneously, handling the relevant tokens according to various definitions of relevance. Following the method by LeCun et al. [25], after the decoder, the output undergoes max pooling to reduce its spatial dimension, with this operation applied along the temporal axis of the frame sequence. Subsequently, as per Srivastava et al. [26], the max pooling output is processed through a dropout layer for regularization and to prevent overfitting during the network's training. Finally, the processed output is sent to a densely connected feed-forward layer, with a number of neurons equivalent to the number of classes in the classification problem. This layer applies a linear transformation followed by a softmax activation function, proposed by Bridle [27], to compute the classification probabilities for each class $\text{attention}_{ij}(Q, K, V)$.

The computation of attention for all tokens can be expressed as the computation of a large matrix using the function:

$$\text{attention}_{ij}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where softmax is taken on the horizontal axis and T means transpose.

Experimental Verification

The experiments were conducted on a dedicated system with the following characteristics: Intel (R) Xeon (R) Gold 6126 CPU at 2.6 64 KiB BIOS, 64 GiB DIMM DDR4 System Memory, 2×GV100GL (Tesla V100 PCIe 32 GB). The proposed framework was developed using Python and the Keras package with Tensorflow. In conducting our experiments, we adopted a two-phase approach. The initial phase assigned relatively less significance to the multiclass classification analysis, with the objective of distinguishing among six different scenarios to ascertain the correctness of each execution. We then progressed to a phase of greater importance and accuracy, concentrating on a binary classification analysis. This crucial phase was specifically designed to evaluate the health status of the subject, determining whether they were in good health or exhibited any health issues.

In the shot boundary detection setup phase of our study, we set specific parameters to optimize the accurate identification of relevant scenes. These choices are the result of extensive testing aimed at precisely capturing the six example cases performed by users:

- Lower motion threshold (-min-percent): set at 1.0%. This lower threshold allows for the detection of scene changes even with minimal movements, ensuring that no significant details in the patients' actions are overlooked.
- Upper motion threshold (-max-percent): set at 10.0%. This limit ensures that excessively large or sudden movements are not mistakenly interpreted as scene changes, maintaining focus on the patients' relevant actions.
- Warm-up period (-warmup): established at 200 frames. This initial phase allows the system to adapt to the video context before starting the actual detection, enhancing accuracy in distinguishing key actions of the patients.

The application of these parameters has proven effective in precisely discerning scene changes that correspond to the specific exercises performed by the patients, providing valuable support in analyzing their activities.

Performance of the Proposed System

To measure the performance of the system, we denote by P a positive condition and by N a negative condition and define the following: TP indicates the number of correctly identified scenes video, FP indicates the number of differently identified scenes video, TN denotes incorrectly identified scenes video, and FN denotes unidentified scenes videos or anomalies.

The performance of the proposed system was evaluated using the following:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Processed Output Evaluation

The classification output, crucial in determining user outcomes across different case studies, was finalized at the end of our algorithm's processing sequence.

Table 2 provides a sample of the output structure, where the "Label" columns indicate the class, each user is assigned to, and "None" is marked in instances where the classification confidence is below 40%.

Table 2 Example of first output of shot boundary transformer neural network classification

Input	Start time	Label	Probability
1	00:00:02	PENTAGON	95.01%
1	00:00:11	PENTAGON_ERRONEOUS	98.28%
1	00:00:16	PENTAGON_ERRONEOUS	98.47%
1	00:00:24	None	35.30%
1	00:00:26	PENTAGON_ERRORNEOUS	99.00%
2	00:00:32	PENTAGON	95.02%
2	00:00:38	HOUSE_ERRONEOUS	98.60%
2	00:00:46	HOUSE_ERRONEOUS	98.76%
2	00:00:52	HOUSE_ERRONEOUS	98.35%
2	00:00:54	HOUSE	94.06%

The process consists of the following three steps:

1. Categorical analysis with threshold criterion: For the label column (Label in Table 2), rather than calculating an average, which is inherently inapplicable to non-numeric data, we identify the mode (the label that occurs most frequently) within a window of 5 consecutive samples. However, to reinforce the statistical significance of this prevalent label, we introduce a probability threshold criterion. A label is considered representative only if the average probability associated with the corresponding samples exceeds 97%.
2. Mode definition: We define the mode as the label that appears most frequently within a window of five samples.
3. Definition of the moving average for probabilities: For the Probability column in Table 2, the moving average (MA) is calculated as the arithmetic mean of the probability values within a window of five samples (video subsequences; Fig. 6).

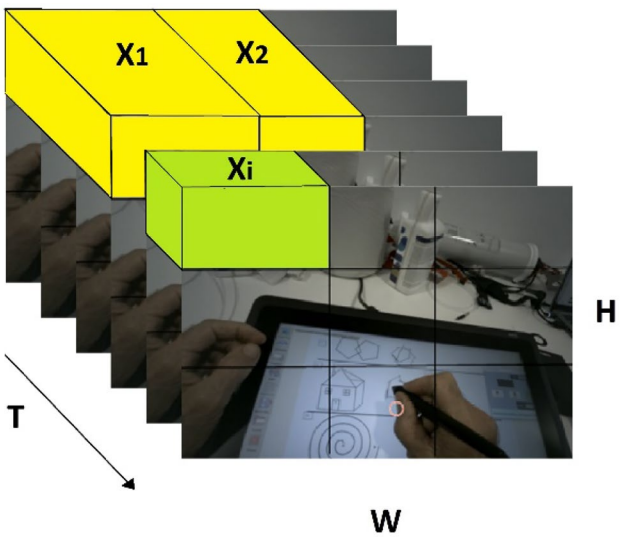

Fig. 6 Sub-sequence

Table 3 Example of processed output from example represented in Table 2

Output	Label
1	PENTAGON_ERRONEOUS
2	HOUSE_ERRONEOUS

For example, applying the method to the first five samples of a time series shown in Table 2 (input 1), we first identify the mode of the labels and calculate the associated probability mean. If the mode occurs at least three times and the average probability exceeds 97%, then this label becomes the representative label for the entire sample window (PENTAGON_ERRONEOUS in this case). Additionally, we calculate the moving average for the probability values.

Our approach allows for a significant summary of the data, associating with each time window a significant categorical label and a numerical estimate of the trend. This methodology enhances the understanding of temporal dynamics in mixed time series (categorical and continuous) and can be generalized to windows of any size and variable probability thresholds.

Table 3 shows the processed output corresponding to Table 2 inputs.

Multiclass Results

For the training data, we considered 80 videos of users who were unable to correctly perform the pentagon test and 102 videos of users who were able to perform the test correctly. By decomposing the videos using shot boundary detection, we obtained 2100 videos. Some of the videos correspond to the explanation time of the database acquisition supervisor, and we also obtained very short videos with a duration on the order of milliseconds of the same scene, which were discarded, as we aimed to maintain a minimum threshold of two seconds of duration for each video.

The discarded short videos are as follows: pentagon (29 videos), pentagon error (27); house (28), house error (31); vertical spiral (29), vertical spiral error (27); line (32), line error (31); spiral (31), spiral error (28); concentric circles (28), consecutive circle error (22).

Furthermore, we duplicated the following videos to avoid an unbalanced dataset: pentagon error (5); vertical spiral error (10); line (15), line error (15); house (12); spiral error (28); circle error (18); pentagon (15).

In total, we acquired a total of 1875 videos that were used for training the network, comprising the following: pentagon (170 videos), pentagon error (170); house (225), house error (336); vertical spiral (105), vertical spiral error (124); line (135), line error (129); spiral (127), spiral error (135); circle (107), circle error (112).

Tables 4 and 5 present the parameters of the final model configuration, which were selected after extensive testing. We will discuss the performance outcomes of these tests in the "Evaluation of the Proposed Framework for Multiclass Problem" section, which focuses on the evaluation of the proposed framework.

In this part, we attempt to automatically detect the class of the video in a 12-class problem. For the Transformer model, we conducted tests over multiple epochs to achieve the highest accuracy rates. The optimal performance was reached at the 150th epoch.

Table 4 Model architecture

Layer	Output shape	Param
Input_12 (input layer)	(MAX_SEQ_LENGTH, NUM_FEATURES)	0
Positional embedding	(MAX_SEQ_LENGTH, NUM_FEATURES)	61,720
Transformer_layer encoder	(MAX_SEQ_LENGTH, NUM_FEATURES)	4,211,716
GlobalMaxPooling1d_5	(NUM_FEATURES)	0
Dropout_5	(NUM_FEATURES)	0
Dense_17	(NUM_CLASSES)	12,300

Table 5 Model hyperparameter

Hyperparameter	Value	Description
MAX_SEQ_LENGTH	60	Maximum length of the input sequence that the model can process
NUM_FEATURES	1024	Number of features per time step or spatial area
IMG_SIZE	128	Dimension of the input images in pixels (128×128)
EPOCHS	120	Total number of complete training cycles on the training data
NUM_CLASSES	12	Total number of complete training cycles on the training data

Evaluation of the Proposed Framework for Multiclass Problem

Following the collection and division of the dataset, 80% was used for training and 20% was set aside for testing, leading our proposed system to achieve good results in classification. The training portion contained 1500 samples, and the testing portion was comprised of 375 samples, reflecting a diverse spectrum of cases, including individuals with cognitive impairment and healthy subjects.

In Table 6, we can observe the values we have obtained as the epochs varied.

Key observations from the data include:

1. Overall trend:
   - There is a discernible trend of precision improvement for many classes as epochs increase, especially noted in "LINE", "LINE_ERRONEOUS", "VERTICAL_SPIRAL", and "VERTICAL_SPIRAL_ERRONEOUS", suggesting enhanced prediction confidence with extended training.
   - Anomaly at 100 epochs, particularly for "SPIRAL" and "VERTICAL_SPIRAL", indicates potential overfitting or ineffective learning past a certain training threshold.
   - By 150 epochs, a majority of the classes exhibit elevated F1-scores compared to earlier epochs, indicative of a balanced enhancement in both precision and recall.
2. Optimal results at 150 epochs:
   - The model appears to strike an optimal balance between precision and recall at 150 epochs, as reflected by high F1-scores. For example, the "LINE" and "LINE_ERRONEOUS" classes achieve impressive precision and recall, culminating in high F1-scores, which means the model is both precise and reliable for these classifications.
   - The "PENTAGON" class demonstrates a notable rise in recall between 120 and 150 epochs, suggesting improved capability in identifying all relevant instances at the latter epoch.
   - It is noteworthy, however, that some classes deviate from this trend, like "CIRCLE_ERRONEOUS", which sees a precision dip from 120 to 150 epochs, potentially due to the model's conservative bias or missing true positives in its quest to minimize false positives.
3. Class-specific performance:
   - "HOUSE_ERRONEOUS" class maintains consistently high and stable scores across epochs, indicating reliable erroneous drawing realization detection for this category.
   - "LINE" class showcases 100% precision at 50 and 120 epochs, representing an ideal scenario, yet its recall is less than perfect, suggesting that while the model's predictions are highly accurate, it doesn't consistently detect all instances of "LINE".
4. Support:
   - Support refers to the number of occurrences of each class in the dataset. In a classification problem, each class has a corresponding support value, indicating how many instances of that class exist in the dataset. Support is particularly relevant in multi-class classification scenarios, helping to understand the distribution of classes and their representation in the dataset. The "support" figures remain unchanged, ensuring a consistent dataset size for each class and a fair comparison across epochs.
5. Areas of concern:
   - The zero precision and recall for the "LINE" class at 100 epochs raise concerns, possibly pointing to a training anomaly or indicating a total detection failure at this stage.

In summary, the choice of 150 epochs is vindicated as it typically delivers the highest F1 scores, denoting an ideal balance between precision and recall. The model's overall accuracy across varying epochs is depicted in Table 7. Nevertheless, the nuanced learning behavior of certain classes at specific epoch milestones underscores the complexity of the model's learning dynamics, necessitating careful consideration when determining the point at which to halt training.

Binary Classification Results

In this section, we attempt to identify the presence or absence of mild cognitive impairment. To maintain an equitable distribution in the binary classification task, the dataset was clustered in a different way. The database initially consisted of 1875 videos, categorized as follows:

- Pentagon (170 videos)
- House (225)
- Vertical Spiral (105)
- Line (135)
- Spiral (127)
- Circle (107)
- Pentagon erroneous (170)
- House erroneous (336)
- Vertical spiral erroneous (124)
- Line erroneous (129)
- Spiral erroneous (135)
- Circle erroneous (112)

Table 6 Processed output

Classes	Precision	Recall	F1-score	Support
20 epochs				
CIRCLE	0.86	0.86	0.86	21
CIRCLE_ERRONEOUS	0.30	0.73	0.43	22
HOUSE	0.74	0.51	0.61	45
HOUSE_ERRONEOUS	0.74	0.76	0.75	67
LINE	0.93	0.56	0.70	25
LINE_ERRONEOUS	0.94	0.59	0.73	27
PENTAGON	0.77	0.79	0.78	34
PENTAGON_ERRONEOUS	0.51	0.74	0.60	34
SPIRAL	0.77	0.74	0.75	27
SPIRAL_ERRONEOUS	0.51	0.67	0.58	27
VERTICAL_SPIRAL	0.93	0.67	0.78	21
VERTICAL_SPIRAL_ERRONEOUS	0.89	0.32	0.47	25
50 epochs				
CIRCLE	1.00	0.81	0.89	21
CIRCLE_ERRONEOUS	0.48	0.73	0.58	22
HOUSE	0.41	0.78	0.53	45
HOUSE_ERRONEOUS	0.65	0.48	0.55	67
LINE	1.00	0.72	0.84	25
LINE_ERRONEOUS	0.89	0.59	0.71	27
PENTAGON	0.82	0.68	0.74	34
PENTAGON_ERRONEOUS	0.67	0.53	0.59	34
SPIRAL	0.75	0.78	0.76	27
SPIRAL_ERRONEOUS	0.46	0.63	0.53	27
VERTICAL_SPIRAL	1.00	0.81	0.89	21
VERTICAL_SPIRAL_ERRONEOUS	0.88	0.60	0.71	25
100 epochs				
CIRCLE	1.00	0.71	0.83	21
CIRCLE_ERRONEOUS	0.42	0.64	0.51	22
HOUSE	0.43	0.51	0.47	45
HOUSE_ERRONEOUS	0.58	0.73	0.64	67
LINE	0.00	0.00	0.00	25
LINE_ERRONEOUS	0.50	0.63	0.56	27
PENTAGON	0.82	0.68	0.74	34
PENTAGON_ERRONEOUS	0.39	0.71	0.50	34
SPIRAL	0.89	0.30	0.44	27
SPIRAL_ERRONEOUS	0.46	0.67	0.55	27
VERTICAL_SPIRAL	1.00	0.33	0.50	21
VERTICAL_SPIRAL_ERRONEOUS	0.70	0.28	0.40	25
120 epochs				
CIRCLE	0.95	0.90	0.93	21
CIRCLE_ERRONEOUS	0.57	0.55	0.56	22
HOUSE	0.52	0.62	0.57	45
HOUSE_ERRONEOUS	0.67	0.72	0.69	67
LINE	1.00	0.80	0.89	25
LINE_ERRONEOUS	0.84	0.59	0.70	27
PENTAGON	0.71	0.59	0.65	34
PENTAGON_ERRONEOUS	0.59	0.56	0.58	34
SPIRAL	0.87	0.74	0.80	27
SPIRAL_ERRONEOUS	0.51	0.78	0.62	27
VERTICAL_SPIRAL	0.90	0.86	0.88	21

Table 6 (continued)

Classes	Precision	Recall	F1-score	Support
VERTICAL_SPIRAL_ERRONEOUS	0.80	0.80	0.80	25
150 epochs				
CIRCLE	0.90	0.86	0.88	21
CIRCLE_ERRONEOUS	0.71	0.45	0.56	22
HOUSE	0.68	0.67	0.67	45
HOUSE_ERRONEOUS	0.76	0.75	0.75	67
LINE	0.96	0.88	0.92	25
LINE_ERRONEOUS	0.92	0.85	0.88	27
PENTAGON	0.69	0.91	0.78	34
PENTAGON_ERRONEOUS	0.67	0.65	0.66	34
SPIRAL	0.88	0.78	0.82	27
SPIRAL_ERRONEOUS	0.55	0.85	0.67	27
VERTICAL_SPIRAL	0.95	0.86	0.90	21
VERTICAL_SPIRAL_ERRONEOUS	0.95	0.76	0.84	25

Table 7 Model accuracy

Epochs	Accuracy	Support
20	67%	375
50	65%	375
100	55%	375
120	70%	375
150	**77%**	**375**

Optimal value is represented in bold letters

These categories were grouped into two main classes: healthy (869) and unhealthy (1006), with the healthy class representing tasks performed correctly by users and the unhealthy class including tasks performed erroneously. To balance the dataset, redundant videos that could bias towards one category over another were excluded, and we removed to prevent an imbalance: house (18), line (6), pentagon erroneous (4), house erroneous (125), vertical spiral erroneous (19), spiral erroneous (8), circle erroneous (5).

We achieved a balanced dataset consisting of 845 videos in the healthy category and 845 videos in the unhealthy category.

In this section, Tables 8 and 9 present the parameters of the final model configuration, which were selected after extensive testing. We will discuss the performance outcomes of these tests in the "Evaluation of the Proposed Framework for Binary Classification" section.

Evaluation of the Proposed Framework for Binary Classification

Following the collection and division of the dataset, 80% was used for training and 20% was set aside for testing, leading our proposed system to achieve good results in classification. The training set contained 1352 samples, and the testing set was comprised of 338 samples, reflecting a diverse spectrum of cases, including individuals with cognitive impairment and healthy subjects.

In Table 10, we can observe the values we have obtained as the epochs varied.

Based on the Table 10, we can obtain several conclusions. It appears that the models trained for 200 and 500 epochs have achieved the best outcomes.

For 200 epochs, the model shows a balance of precision and recall, with the "healthy" class achieving an F1-score of 0.80 and the "not healthy" class at 0.84, indicating a strong performance in both the positive predictive value and the sensitivity of the test.

Table 8 Model architecture

Layer	Output shape	Param
Input_12 (input layer)	(MAX_SEQ_LENGTH, NUM_FEATURES)	0
Positional embedding	(MAX_SEQ_LENGTH, NUM_FEATURES)	61,440
Transformer_layer encoder	(MAX_SEQ_LENGTH, NUM_FEATURES)	4,211,716
GlobalMaxPooling1d_5	(NUM_FEATURES)	0
Dropout_5	(NUM_FEATURES)	0
Dense_17	(NUM_CLASSES)	2050

Table 9 Model hyperparameter

Hyperparameter	Value	Description
MAX_SEQ_LENGTH	60	Maximum length of the input sequence that the model can process
NUM_FEATURES	1024	Number of features per time step or spatial area
IMG_SIZE	128	Dimension of the input images in pixels (128 × 128)
EPOCHS	120	Total number of complete training cycles on the training data
NUM_CLASSES	12	Total number of complete training cycles on the training data

Table 10 Processed output

Classes	Precision	Recall	F1-score	Support
150 epochs				
HEALTHY	0.78	0.74	0.76	150
NOT_HEALTHY	0.78	0.82	0.80	171
200 epochs				
HEALTHY	0.84	0.77	0.80	150
NOT_HEALTHY	0.81	0.87	0.84	171
500 epochs				
HEALTHY	0.86	0.75	0.80	150
NOT_HEALTHY	0.80	0.89	0.84	171
800 epochs				
HEALTHY	0.70	0.72	0.71	150
NOT_HEALTHY	0.75	0.73	0.74	171

When the model is trained for 500 epochs, we observe a similar performance, with a slight variation in the F1-scores: the "healthy" class maintains an F1-score of 0.80, while the "not healthy" class also remains at 0.84. This suggests that increasing the number of epochs to 500 offers little to no significant improvement in the model's ability to generalize, although it may require slightly more computational time.

At 800 epochs, there is a noticeable decline in performance across all metrics for both classes when compared to 200 and 500 epochs. The "healthy" class drops to an F1-score of 0.71 and the "not healthy" class to 0.74, indicating that too many epochs may lead to overfitting or other forms of performance degradation.

In summary, training for 200 epochs appears to be the most efficient in terms of achieving high predictive performance without unnecessary computational expense. Training for 500 epochs does not show a significant difference in performance, suggesting that the additional epochs may not be worth the extra time required. It is also clear that training beyond this point, such as at 800 epochs, does not improve and may even degradate the model's performance.

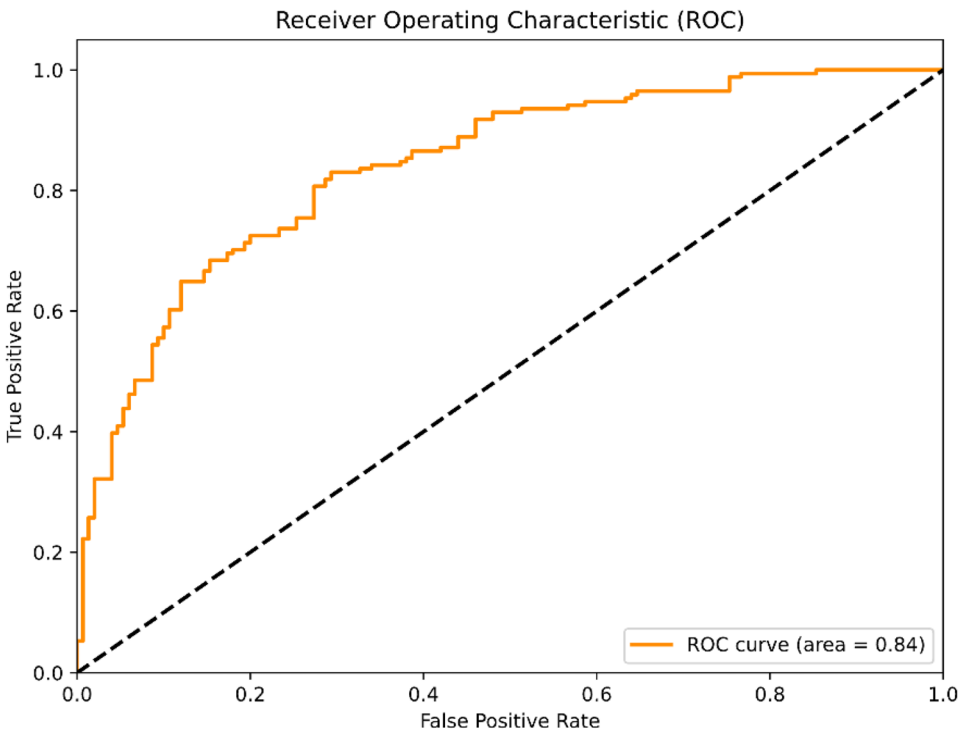

Fig. 7 ROC figure of binary classification

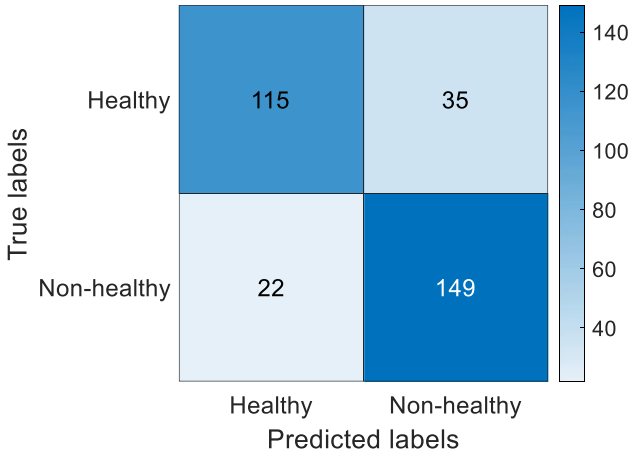

Fig. 8 Confusion matrix for two-class problem

Figure 7 depicts the model's ROC curve. The plotted ROC curve serves as a good indicator of the model's ability to differentiate between the two classes.

Some performance indicators are:

- Area under the curve (AUC): The AUC is 0.84, which denotes a robust discriminative capacity of the model. Generally, an AUC ranging from 0.8 to 0.9 is considered very good, suggesting that the model is well-calibrated and can distinguish between "healthy" and "not healthy" subjects with a high degree of probability.
- Trade-off between TPR and FPR: The curve illustrates that as the true positive rate increases (more "healthy" subjects correctly identified), the false positive rate (more "not healthy" subjects incorrectly identified as "healthy") also increases, which is typical in binary classification settings. The goal is to maximize the TPR while minimizing the FPR.
- Model performance: For the most part, the curve lies well above the dashed diagonal line that represents random classification (AUC = 0.5). This indicates that the model's performance is significantly better than chance.
- Optimal threshold: The specific point along the curve considered "optimal" will depend on the application context and the desired balance between TPR and FPR. For instance, in some medical applications, it might be preferable to minimize false negatives (thereby maximizing sensitivity), even at the expense of accepting more false positives.

Table 11 Model accuracy

Epochs	Accuracy	Support
150	78%	321
200	82%	321
500	**83%**	**321**
800	72%	321

Optimal value is represented in bold letters

In conclusion, the model appears to be effective at distinguishing between "healthy " and "not healthy" subjects with a commendable level of accuracy. However, the choice of threshold for classifying subjects as "healthy " or "not healthy" should be guided by clinical considerations that weigh the importance of correctly identifying "healthy" cases against the risk of false alarms.

Figure 8 depicts the confusion matrix, which summarizes the main performance ([28], page 438):

- True negatives (TN): The model has correctly identified 115 cases as "healthy". This indicates that the model is fairly reliable in recognizing the absence of "non-healthy" conditions.
- False positives (FP): There were 35 instances where the model incorrectly classified "healthy" cases as "non-healthy". Such errors can be problematic in medical contexts, as they may lead to unnecessary further testing or treatments.
- False negatives (FN): The model failed to identify 22 cases of "non-healthy", erroneously classifying them as "healthy". These errors are often considered more serious in clinical contexts, as they imply a missed treatment of a condition that requires attention.
- True positives (TP): With 149 cases correctly identified as "non-healthy", the model demonstrates a good ability to detect conditions when they are indeed present.

In Table 11, we summarize the best values obtained; in particular, the table outlines the accuracy of a model at various epochs—150, 200, 500, and 800 while maintaining a constant support of 321 cases. This pattern suggests that the model benefits from additional training up to a point (500

Table 12 Comparison of model accuracy in existing literature and our work

Reference	Methodology	Accuracy
Kruthika et al. [30]	Gaussian Naive Bayes	93%
Kruthika et al. [30]	K-nearest neighbor (KNN)	93%
Kruthika et al. [30]	Support vector machine (SVM)	93%
Kruthika et al. [30]	SVM + KNN + PSO	93%
Liu et al. [31]	Cascaded CNNs for AD diagnosis	93%
Payan et al. [32]	DNN with sparse AE and CNN	89%
Sarraf et al. [33]	CNN MRI	98%
Sarraf et al. [33]	CNN fMRI	99%
Erdas et al. [29]	3DCNN	96%
Erdas et al. [29]	ConvLSTM	95%
This work	**ShotBoundary Transformer**	**83%**

epochs), but beyond this, its performance declines, which may indicate overfitting. This occurs when the model learns the training data too well, including noise and outliers, leading to a decrease in its ability to generalize to new data.

Discussion and Conclusions

In this section, we discuss the experimental results obtained, comparing them with the state of the art and current techniques in the recognition of neurodegenerative diseases showed in Table 12. Erdaş et al., in their study Erdas et al. [29], introduced an innovative method using ConvLSTM and 3D CNN to analyze unidimensional human gait data, converting it into bidimensional and tridimensional formats for analysis. They describe a pioneering approach that converts gait data into two-dimensional QR codes, subsequently used to feed ConvLSTM and 3D CNN models. The experiment showed promising results in distinguishing between subjects with neurodegenerative diseases (NDD) and healthy controls. However, the outstanding results raise methodological questions: precise and consistent data collection from the ground reaction force (GRF) sensor is essential to maintain result reliability, and QR code conversion might introduce errors or information loss. Moreover, individual variability in the dataset could make it challenging to distinguish specific gait patterns for NDD.

Previous studies, like Sarraf et al. [33], have used CNNs to distinguish between HC and AD in older adults through the extraction of scale- and position-invariant features. They proposed two workflows, one with structural fMRI data and the other with structural MRI data, both presupposing a prediagnosis of Alzheimer's, limiting the applicability of such methods for early disease detection. In contrast, our methodology could be employed prior to resorting to fMRI or MRI.

Payan et al. [32] employ deep learning to identify Alzheimer's from MRI images, combining DNN, SAE, and CNN. The model, applied to 3D MRI images of seniors over 75 years, achieved an accuracy of 89.47% in classification, highlighting its effectiveness in complex data analysis. However, it still relies on MRI, not anticipating neurodegenerative symptoms.

Liu et al. [31] use machine learning, specifically CNNs, for Alzheimer's diagnosis through neuroimaging. They constructed cascading convolutional neural networks to analyze multimodal neurological images, particularly MRI and PET, from Weiner et al.'s ADNI data. This study marked a progression in AI's use in medical image analysis, especially for AD diagnosis, although their method focuses on confirming diseases rather than early detection.

Kruthika et al. [30] combined Gaussian Naive Bayes, KNN, and SVM to analyze MRI scans, focusing on brain volume and thickness measures. Despite limitations in discriminating between MCI states and cognitive normality, their study underlines the need for more advanced biomarkers.

Although presented references in Table 12 outperform our system, we have to take into account that straight comparison is not possible due to the following aspects:

- Existing literature relies on more sophisticated signals such as human gait, MRI, fMRI, and PET. These signals require an expensive hardware, while our system relies on a simple and cheap frontal camera mounted on the head.
- Experimental results of existing literature is based on homogeneous databases, where there is a single pathology and the ground truth was obtained after a medical diagnose. In our case, we probably have a set of diverse pathologies; medical diagnose is not available, and ground truth is derived on a single test, which is the pentagon copying test.

Our methodology proves to be promising in the early detection of cognitive impairment disorders. The decision to use the Transformer network was crucial: unlike CNNs, which analyze individual frames, the Transformer has enabled us to process sequential data. This is beneficial because a single frame might not be indicative; for example, if a patient momentarily diverts their gaze during a test, it could lead to a misdiagnosis. However, by evaluating the entire data sequence, the Transformer network yields a more complete and precise analysis. While the limited number of samples has impacted our outcomes, our study marks a significant advancement.

Usually, the cost of wearable eye-tracker systems is quite large (4000–24,000 euros); in this paper, we presented a new assessment methodology for cognitive impairment based on a frontal camera, which is a simpler and cheaper hardware. The proposed system ensures the detection and diagnosis of users affected by the stated problems. This phase could precede a series of tests [34] such as clinical, blood, urine, or spinal cord examinations, neuropsychological tests to measure memory, problem-solving ability, degree of attention, and the ability to converse and count; brain CT scans may also be conducted for identification of any possible signs of abnormality, most of which almost invariably turn out to be invasive examinations as well. These examinations allow the physician to excise other possible causes. In neurodegenerative diseases, early diagnosis is very important both because it offers the possibility of treating some symptoms of the disease and because it allows the patient to plan for his or her future while still being able to make decisions. With the proposed method, we can make an early diagnosis of cognitive functioning in a non-invasive way and then address any clinical testing for ascertainment; at the end

of all processes of elaboration, we have found an accuracy of 83% in a two-class problem and 77% in a multi-class problem. Moreover, for the first experimental phase of verification and operation of the proposed system, we have limited ourselves to the number of training data. The proposed technique may open new frontiers in the clinical and early detection fields. We set as a future goal the extension of the dataset and testing on a large scale. In addition, the tests were carried out on elderly patients; the goal will be to diagnose neurodegenerative diseases still in the process of development much earlier on young patients.

In the interpretation of our study results, it is essential to acknowledge certain limitations that may impact the generalizability and robustness of our findings:

- We did not systematically record the medications being taken by participants, and it is recognized that pharmaceutical interventions can influence cognitive performance and may have introduced variability into our results. Additionally, the absence of recorded information pertaining to participants' previous medical history, such as a history of stroke, introduces a potential confounding factor that was not accounted for in our analysis.
- A subset of individuals within our study exhibited difficulties with writing, potentially attributable to a deficit in their educational background.
- The use of tablets for tasks involving handwriting may be unfamiliar to elderly individuals, potentially influencing their performance on such devices.

It is imperative to acknowledge that certain limitations in our study arise from deliberate choices made to prioritize participant privacy. Specifically, the decision not to record participants' medication usage and detailed medical history was made to uphold ethical considerations and mitigate potential privacy concerns. The sensitive nature of health-related information necessitated a cautious approach to data collection, particularly given the potential impact on individual privacy and confidentiality. Future investigations may explore alternative methodologies that address these limitations while maintaining the highest standards of participant privacy.

Funding This work has been supported by Grant PID2020-113242RB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER EU and European project FEDER Territorial Competitiveness Specialization Project of Mataro-Maresme (PRE/161/2019).

Data Availability The datasets generated during the current study are available from the corresponding author upon a reasonable request.

Declarations

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study, formal consent is not required.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare no competing interests.

References

1. Ding Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18f-fdg pet of the brain. Radiology. 2019;290:456–64. https://doi.org/10.1148/radiol.2018180958.
2. Doulamis N, Voulodimos A. Fast-mdl: fast adaptive supervised training of multi-layered deep learning models for consistent object tracking and classification. 2016 IEEE International Conference on Imaging Systems and Techniques (IST). 2016;318–323.
3. He Z, Dieciuc M, Carr D, et al. New opportunities for the early detection and treatment of cognitive decline: adherence challenges and the promise of smart and person-centered technologies. BMC Digit Health. 2023;1:7. https://doi.org/10.1186/s44247-023-00008-1.
4. Liss JL, Seleri Assunção S, Cummings J, Atri A, Geldmacher DS, Candela SF, Devanand DP, Fillit HM, Susman J, Mintzer J, Bittner T, Brunton SA, Kerwin DR, Jackson WC, Small GW, Grossberg GT, Clevenger CK, Cotter V, Stefanacci R, Wise-Brown A, Sabbagh MN. Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. J Intern Med. 2021;290(2):310–334. https://doi.org/10.1111/joim.13244. Epub 2021 Mar 31. PMID: 33458891; PMCID: PMC8359937.
5. Tobii pro glasses 3. 2023. https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3#video.
6. Faundez-Zanuy M, Fierrez J, Ferrer MA, et al. Handwriting biometrics: applications and future trends in e-security and e-health. Cogn Comput. 2020;12:940–53. https://doi.org/10.1007/s12559-020-09755-z.
7. Mucha J. HandAQUS Handwriting Acquisition Software - user manual. 2021. https://doi.org/10.13140/RG.2.2.16562.53440. available at Github handAQUS. https://github.com/BDALab/HandAQUS.
8. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
9. Budson AE, Solomon PR. Chapter 2 - evaluating the patient with memory loss or dementia, Editor(s): Andrew E. Budson, Paul R. Solomon, Memory loss, Alzheimer's disease, and dementia (Second

Edition), Elsevier, 2016, Pages 5–38, ISBN 9780323286619. https://doi.org/10.1016/B978-0-323-28661-9.00002-0.
10. Faundez-Zanuy M, Mekyska J, Impedovo D. Online handwriting, signature and touch dynamics: tasks and potential applications in the field of security and health. Cogn Comput. 2021;13:1406–21.
11. Ghadiri-Sani M, Larner AJ. Head turning sign. J R Coll Physicians Edinb. 2019;49(4):323–6. https://doi.org/10.4997/JRCPE.2019.416. PMID: 31808463.
12. Durães J, Tábuas-Pereira M, Araújo R, Duro D, Baldeiras I, Santiago B, Santana I. The head turning sign in dementia and mild cognitive impairment: its relationship to cognition, behavior, and cerebrospinal fluid biomarkers. Dement Geriatr Cogn Disord. 2018;46(1–2):42–9. https://doi.org/10.1159/000486531. Epub 2018 Aug 9 PMID: 30092564.
13. Yesavage JA, BrinK TL, Rose TL, Lum O. Development and validation of a geriatric depression scale: a preliminary report. J Psychiat Res. 1983;17(1):37–49.
14. Opwonya J, Ku B, Lee KH, Kim JIl, Kim JU. Eye movement changes as an indicator of mild cognitive impairment. Front Neurosci. 2023;17. https://www.frontiersin.org/articles/https://doi.org/10.3389/fnins.2023.1171417.
15. Howard CW. Neural networks for cognitive testing: cognitive test drawing classification. Intell-Based Med. 2023;8:100104, ISSN 2666–5212. https://doi.org/10.1016/j.ibmed.2023.100104.
16. Gao X, Alimoradi S, Chen J, Hu Y, Tang S. Assistance from the ambient intelligence: cyber–physical system applications in smart buildings for cognitively declined occupants. Eng Appl Artif Intell. 2023;123:106431, ISSN 0952-1976. https://doi.org/10.1016/j.engappai.2023.106431.
17. Cossu G, Colosimo C. Hyperkinetic movement disorder emergencies. Curr Neurol Neurosci Rep. 2017;17:6. https://doi.org/10.1007/s11910-017-0712-7.
18. Roze E, Coelho-Braga MC, Gayraud D, Legrand AP, Trocello J-M, Fenelon G, Cochen V, Patte N, Viallet F, Vidailhet M, et al. Head tremor in Parkinson's disease. Movement disorders: official journal of the Movement Disorder Society. 2006;21:1245–8.
19. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9. https://doi.org/10.1111/j.1532-5415.2005.53221.x.Erratum.In:JAmGeriatrSoc.2019Sep;67(9):1991. PMID: 15817019.
20. Nagaratnam N, Nagaratnam K, O'Mara D. Intersecting pentagon copying and clock drawing test in mild and moderate Alzheimer's disease. J Clin Gerontol Geriatrics. 2014;5(2):47–52, ISSN 2210-8335. https://doi.org/10.1016/j.jcgg.2013.11.001.
21. Candela F, et al. Shot boundary detection and convolutional neural network for video classification. 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE 2023.
22. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
23. Huang G, Liu Z, Van Der ML., Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;4700–4708.
24. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. 2009;248–255.
25. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86:2278–324.
26. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014;15:1929–58.
27. Bridle JS. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. Neurocomputing: Algorithms, architectures and applications. 1990;227–236.
28. Smékal Z, Sklenář J. 1D and 2D analog, discrete, and digital signal processing. první. první. Brno, Czech Republic: Brno University of Technology - VUTIUM Press. 2023;454. ISBN: 978-80-214-6143-7.
29. Erdaş ÇB, Emre S, Seda K. Neurodegenerative disease detection and severity prediction using deep learning approaches. Biomed Signal Process Control. 2021;70:103069, ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2021.103069.
30. Kruthika KR, Maheshappa HDR. Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. Inform Med Unlocked. 2019;14:34–42.
31. Liu M, Cheng D, Wang K. Alzheimer's Disease Neuroimaging Initiative. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics. 2018;16(3–4):295–308. https://doi.org/10.1007/s12021-018-9370-4. PMID: 29572601.
32. Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. ICPRAM. 2015;(2):355–62. SciTePress. https://doi.org/10.48550/arXiv.1502.02506.
33. Sarraf S, Tofighi G, et al. Deepad: Alzheimer's disease classification via deep convolutional neural networks using MRI and FMRI. bioRxiv. 2016. https://doi.org/10.1101/070441.
34. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. 2011;263–269.