

Deep Learning in Mild Cognitive Impairment Diagnosis using Eye Movements and Image Content in Visual Memory Tasks

Tomás Silva Santos Rocha^{1 a}, Anastasiia Mikhailova^{1,2,3 b}, Moreno I. Coco^{4 c},
and José Santos-Victor^{1 d}

¹ Institute for Systems and Robotics, Laboratory for Robotics and Engineering
Systems, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

^a tomasrocha01@tecnico.ulisboa.pt, ^d jose.santos-victor@tecnico.ulisboa.pt

² Department of Psychology, University of Chicago, Chicago, USA

³ Institute for Mind and Biology, University of Chicago, Chicago, USA

^b amikhailova@uchicago.edu

⁴ Department of Psychology, Sapienza Università di Roma, Rome, Italy

^c moreno.coco@uniroma1.it

Abstract. The global prevalence of dementia is projected to double by 2050, highlighting the urgent need for scalable diagnostic tools. This study utilizes digital cognitive tasks with eye-tracking data correlated with memory processes to distinguish between Healthy Controls (HC) and Mild Cognitive Impairment (MCI), a precursor to dementia. A deep learning model based on VTNet was trained using eye-tracking data from 44 participants (24 MCI, 20 HCs) who performed a visual memory task. The model utilizes both time series and spatial data derived from eye-tracking. It was modified to incorporate scan paths, heat maps, and image content. These modifications also enabled testing parameters such as image resolution and task performance, analysing their impact on model performance. The best model, utilizing $700 \times 700px$ resolution heatmaps, achieved 68% sensitivity and 76% specificity. Despite operating under more challenging conditions (e.g., smaller dataset size, shorter task duration, or a less standardized task), the model's performance is comparable to an Alzheimer's study using similar methods (70% sensitivity and 73% specificity). These findings contribute to the development of automated diagnostic tools for MCI. Future work should focus on refining the model and using a standardized long-term visual memory task.

Keywords: Mild Cognitive Impairment · Memory · Eye Movements · Deep Learning · Image Content.

1 Introduction

The WHO's "Global Status Report on the Public Health Response to dementia" [21] highlights dementia as a growing global health issue, affecting over 55 million people, projected to rise to 139 million by 2050, with a disproportionate

impact expected in middle-income countries. This emphasizes the urgent need for diagnostic tools and interventions, particularly as populations age. Dementia is characterized by cognitive decline that interferes with daily life, including memory loss and impaired thinking. Alzheimer’s disease is the most common form, though other types include vascular, frontotemporal, and Lewy body dementia [4]. Before developing dementia, patients go through a stage called Mild Cognitive Impairment (MCI), experiencing mild cognitive changes that do not significantly interfere with daily life activities. Early detection of MCI provides a critical window for interventions that may delay progression to dementia [22]. There are two types: amnesic, which impacts memory and has a higher risk of Alzheimer’s, and non-amnesic, which affects attention, language, or visuospatial skills and may indicate risk for other dementias, such as frontotemporal or vascular dementia [25]. MCI may also involve single-domain or multiple-domain impairments, with the latter suggesting more significant progression to dementia [25, 10].

1.1 Diagnosing MCI

MCI is influenced by age, genetics (e.g., APOE ϵ 4 allele), lifestyle, and cardiovascular health and is associated with biomarkers such as hippocampal atrophy and amyloid-beta accumulation. However, diagnosing it promptly remains challenging, as neuroimaging and genetic tests are primarily used for the confirmation of the diagnosis, and lifestyle factors lack diagnostic precision [15, 12]. The limited availability of specialists restricts the widespread use of standard cognitive tests like MMSE and MoCA. When combined with delayed symptom recognition, this often results in MCI being detected only when cognitive decline is noticeable, limiting early intervention opportunities [2, 5].

To tackle these challenges, digital assessments on computers, tablets, and smartphones are emerging as practical and cost-effective alternatives [27]. When integrated with wearable sensors, cameras, or eye-tracking devices, these tools capture many relevant data without adding time costs, enhancing their diagnostic potential [14].

Eye-tracking integration in MCI diagnosis has gained attention, with studies suggesting that eye movement patterns could serve as early indicators of cognitive decline [36].

1.2 Eye-movements: A window to the memories

Eye movements, essential for visual perception, offer insight into cognitive processing and memory. Key movements include saccades, rapid gaze shifts between points, and fixations, where the eyes pause to process information. Saccades allow quick repositioning but temporarily suppress visual intake, while fixations enable detailed analysis, with durations varying by task [24]. Fixations are particularly useful, as their longer periods allow for easier detection. Eye movements significantly impact memory mechanisms, aiding encoding and retrieval. Increased fixations and shifts in gaze while encoding a scene improve memory

recall, as eye movements facilitate scene exploration [8]. During encoding, eye movement patterns differ notably between MCIs and healthy individuals. MCI patients often have delayed saccadic responses and less accurate saccade targeting, leading to inefficient visual exploration. Their fixations are longer but fewer, possibly indicating attention and visual processing difficulties. This reduced fixation frequency and longer saccadic latency contribute to limited scene exploration, impacting their ability to recognize and remember visual details [36]. In the literature, we also see that image content influences eye movements by engaging in a process where our eyes selectively focus on elements perceived as necessary, consequently influencing the memory processes [32].

1.3 Deep Learning and MCI

To enhance the digital tools deep learning algorithms are being employed, improving diagnostic accuracy and scalability, thus making the timely detection of MCI a possibility [36].

Algorithms based on various methods have been employed, such as Recursive Neural Networks (RNN), Convolutional Neural Networks (CNN), autoencoders, transformers, and others [30]. The transformer based architectures usually perform better than the others, however they require large amounts of data to be properly trained [30, 16]. One problem with recent studies is the fact that, typically there is no differentiation between MCI patients and patients with more advanced dementias, such as Alzheimer’s, which can bring biases to the results obtained [30].

Building upon recent advancements, this study explores deep learning models that use eye-tracking data collected from MCI patients and Healthy Controls (HCs). The participants perform a visual long-term active memory task to predict MCI, which have been shown to have better diagnostic accuracy [34, 36]. Additionally, this paper investigates incorporating image content, a topic that has not been extensively researched, and memory performance to try to enhance the diagnostic process further.

2 Method

Initially, we planned to use our collected data; however, the process is ongoing, and the sample size is insufficient. Instead, we utilized data from [6], whose experiment closely aligned with our study. This study [6] examined how semantically related images impact memory performance in MCI and HC groups.

2.1 Participants and Procedure

Selection criteria: (a) ages between 50 and 90; (b) minimum of three schooling years; (c) normal or corrected-to-normal vision with no history of eye surgery; (d) no history of neurological (except memory disturbances) or psychiatric disorders; (e) no history of alcohol/substance abuse or medications affecting cognition.

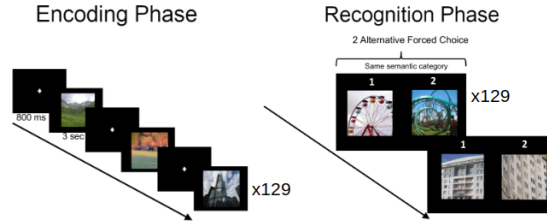


Fig. 1: Schematic of the experiment performed by the participants. Left: Encoding Phase; Right: Recognition Phase. [6]

Data collection resulted in 44 participants (24 MCI patients: 71.92 ± 9.06 years old, 9.83 ± 4.50 schooling years; 20 HCs: 68.50 ± 8.79 years old, 11.05 ± 5.10 schooling years), and a total of 79 sessions after filtering for (a) follow-up participation; (b) performance above chance; (c) successful eye-tracker calibration.

The visual memory task was conducted on a 17-inch monitor (1600×900 resolution, 50 Hz refresh rate). Eye movements were recorded with an EyeTribe eye-tracker (55 Hz sampling), with accuracy validated in [20], and calibrated at a low resolution (55 Hz), which corresponds to a sample every 33 ms. The experiment was implemented using OpenSesame [17] and PyGaze [7] to calibrate the eye-tracking using a 9-point procedure before each session.

Each session, approximately 30 minutes, consisted of one encoding and recognition phase (Fig. 1). During encoding, participants viewed 129 real-life images (700×700 pixels) randomly selected from 10 lists (834 images total), each shown for 3 seconds with an 800 ms interval between trials.

In the recognition phase, participants saw two images side by side, one of which was previously seen, and were asked to identify the previously seen image.

2.2 Data Curation and Processing

Pre-processing addressed missing eye-tracking data from blinks, movements, or detection errors, removing fixations outside the image area [11]. Trials with $\geq 20\%$ missing data or gaps over one second were discarded; otherwise, only lost points were removed [13]. This reduced the dataset from 10,191 (4,773 HCs, 5,418 MCI) to 8,615 trials (4,225 HCs, 4,390 MCI). To mitigate imbalance, data from the best-performing MCI patient was excluded, further refining the dataset to 8,499 trials (4,225 HCs, 4,274 MCI) [3, 19].

The final dataset was divided into 10 balanced folds for a 10-fold nested cross-validation for model evaluation and hyperparameter tuning. Data from each participant was assigned to a single fold to prevent bias. Each fold contained a training set ($\sim 80\%$; 35 participants: 19 MCIs and 16 HCs), validation set and test set (10% each; ~ 5 participants: 3 MCIs and 2 HCs each).

2.3 The Model

We selected the VTNet [31] due to its prior success in Alzheimer’s prediction [33]. Model training and testing were conducted using PyTorch 2.3.0+cu121 [1] with

the NVIDIA® CUDA® Toolkit 12.1 [18], running on an Intel® Core™ i7-9700 CPU, 64 GB RAM, and an NVIDIA® GeForce RTX™ 3090 GPU (24 GB memory). The code is available on GitHub¹.

Architecture VTNet, Fig. 2(a) consists of three key components: an RNN path, a CNN path, and a classifier. The RNN path processes gaze position and pupil size time-series (x, y axes for both eyes) across 107 time steps using a self-attention layer followed by a Gated Recurrent Unit (GRU) to capture sequential patterns. The CNN path processes a visual representation, with two convolutional layers and max pooling, encodes spatial gaze distributions. Outputs from both paths are concatenated and passed through two fully connected layers for trial-wise predictions, which can be aggregated at the participant level.

Proposed Models To refine the base model [33], we explored several adaptations. First, scanpath images ($700 \times 700px$), visualizing gaze sequences, were resized to $256 \times 256px$ to more closely align with the model’s expected dimensions ($230 \times 350px$). We then compared scanpaths with gaze heatmaps, which highlight fixation density using a Multivariate Normal Distribution (Equation 1):

$$f(x, y) = \frac{\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right)}{2\pi\sigma_x\sigma_y} \quad (1)$$

Where μ is the gaze point and σ is a standard deviation of 1° visual angle.

Next, we investigated whether image content could aid MCI prediction by integrating visual representations. Three approaches were tested: (1) a 4D image combining RGB channels with a heatmap mask (Fig. 2(d)), (2) a 2D Gray Scale, GS, image with a separate heatmap channel (Fig. 2(e)), and (3) an RGB image multiplied by the heatmap mask (Fig. 2(f)).

Finally, we examined the impact of task performance on model accuracy using two strategies: one incorporating trial-by-trial performance in the concatenated vector, Single Trial Score (STS) and another combining each participant’s overall score with the model’s average trial performance, Average Test Score (ATS). These two models combine the task performance with the best-performing models from the ones described above.

2.4 Training, Architecture Optimization and Model Analysis

Each model was trained with a batch size 32, balancing computational efficiency and model performance. We employed the Cross-Entropy Loss function and the Adam optimizer. The initial learning rate was set to 10^{-5} , as higher rates resulted in convergence failure. The learning rate was halved after five consecutive validation epochs without loss reduction, down to a minimum of 10^{-6} to facilitate fine-tuning. Early stopping was implemented to prevent overfitting, halting

¹ <https://github.com/tometaro07/eye-tracking-mci.git>

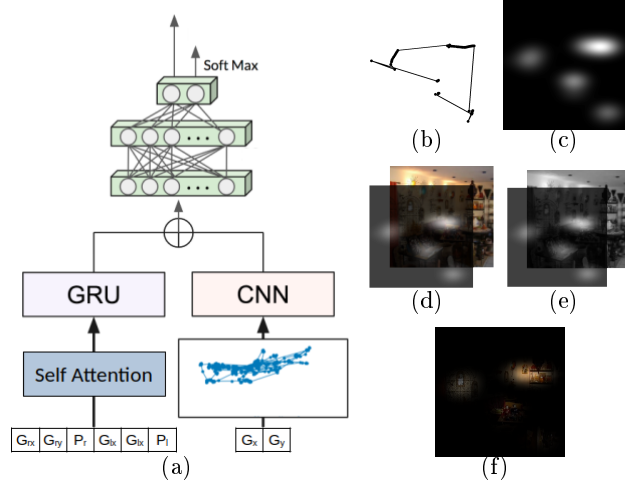


Fig. 2: Architecture for the models described and examples of the different types of inputs used for the CNN pathway. (a) Architecture: Left path - RNN path, with Self-Attention Layer and GRU Layer (G_{rx} and G_{ry} coordinates of the gaze for the right eye; G_{lx} and G_{ly} coordinates of the gaze for the left eye; P_x and P_l pupil size for left and right eye); Left Path - CNN Path (G_x and G_y coordinates of gaze); Green Part - Classifier [31]; (b) Scanpath; (c) Heatmap; (d) 4D Image; (e) GS image and Heatmap; (f) 3D Image multiplied by Heatmap mask

training after 10 epochs without improvement in validation loss and restoring the model to its best-performing weights.

Hyperparameter tuning focused on optimizing the number of convolutional layer channels (6, 8, 12, 16) and the output vector sizes for the RNN and CNN paths (8, 16, 32, 64, 128). The optimal model configuration was determined by maximizing the combined sensitivity and specificity. To contextualize our findings, we compared our results to those obtained in the Image Description Task from [33]. This study, which also utilized VTNet, involved a similar image exploration task but was exclusively applied to Alzheimer’s patients. To provide additional insight into the visual complexity of the input images, we calculated image entropy using the following formula [35]:

$$H = - \sum_{i=0}^{n-1} p_i \log_2 p_i \quad (2)$$

3 Results & discussion

3.1 Scanpaths

We first analysed models similar to the original study [33], using scanpath inputs of $256 \times 256px$ and $700 \times 700px$. Results are presented in Table 1.

Table 1: Mean Sensitivity and specificity, in percentage, with respective standard deviation, for the models using scanpaths and heatmaps as input. The results from the image description task in the original study [33] are also included.

	Scanpath		Heatmap		Image Description Task [33]
	(256 × 256)	(700 × 700)	(256 × 256)	(700 × 700)	
Sensitivity	52.63 ± 37.91	47.37 ± 39.53	63.16 ± 28.50	68.42 ± 27.26	70 ± 0.02
Specificity	70.56 ± 41.08	58.82 ± 41.50	52.94 ± 43.46	76.47 ± 27.64	73 ± 0.02

The $256 \times 256px$ model exhibited lower sensitivity (53%) in differentiating between MCI patients and HCs, compared to the Image Description Task [33], the base study (70%). This difference is likely due to subtler distinctions between MCI and healthy controls (HCs), compared to the differences between dementia and HC classification in the original study. Additionally, the higher standard deviation ($256 \times 256px$ model: $\sim 41\%$; Base study: 0.02%) may reflect the smaller sample size (our dataset: 44 participants; Base study dataset: 144 participants), making misclassification more impactful.

Comparing input sizes, the $700 \times 700px$ model performed worse, likely because scanpath images are 98% empty space, making smaller images more effective at highlighting relevant patterns and preventing overfitting. Moreover, the original architecture was designed with $230 \times 350px$ inputs in mind, reinforcing the importance of choosing an appropriate image size, a trend also observed in handwriting recognition and image segmentation models [29, 26].

3.2 Heatmaps

Contrary to the scanpath models, the $700 \times 700px$ heatmap model outperformed the $256 \times 256px$ one due to the loss of information when resizing heatmaps, which led to a performance drop, as shown in Table 1. Research indicates that increasing image size benefits performance up to a point, after which gains plateau [28].

To compare the best models from the heatmap and scanpath models, the $700 \times 700px$ heatmap model and the $256 \times 256px$ scanpath model, we first have to examine the image entropy of the model’s inputs. Using Fig. 3 as an example, we can see the gaze heatmaps from two different participants (one MCI, Fig. 3(c), and one HC, Fig. 3(b)) viewing the same scene, Fig. 3(a). The entropies of these two heatmaps are 2.65 for the HC and 1.8 for the MCI, indicating a lesser spread of the MCI gaze, and consequently, lower image exploration.

The entropy analysis across trials (summarized in Table 2) revealed that HCs consistently had higher entropy values than MCI patients, highlighting differences in encoding success. Additionally, correct trials exhibited higher entropy, especially in HCs, suggesting greater exploration improves memory encoding.

Scanpath entropies (Table 2) showed minimal differences between classes (0.003), significantly lower than the difference between classes in heatmap entropies (0.109). This suggests that entropy plays a role in model performance, where higher entropy differences between classes correlate to better results, evidence also supported by [23] and [9].

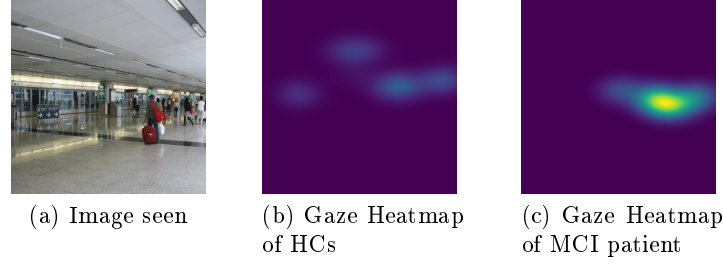


Fig. 3: Examples of the types of inputs used to test if image content can improve the model's result

Table 2: Average entropy values for the $700 \times 700px$ heatmaps and $256 \times 256px$ scanpath, with respective standard deviation. These values are calculated across the participants' group (HCs and MCI patients), and also by grouping the trials in which participants choose the correct answer or not.

	Heatmap $700 \times 700px$			Scanpath $256 \times 256px$		
	HCs	MCI Patients	Total	HCs	MCI Patients	Total
Got Trial	$2.614 \pm$	$2.512 \pm$	$2.567 \pm$	$0.368 \pm$	$0.377 \pm$	$0.372 \pm$
Correct	0.443	0.441	0.445	0.146	0.180	0.163
Got Trial	$2.494 \pm$	$2.415 \pm$	$2.446 \pm$	$0.350 \pm$	$0.347 \pm$	$0.348 \pm$
Wrong	0.476	0.443	0.458	0.157	0.175	0.169
Total	$2.589 \pm$	$2.480 \pm$	$2.534 \pm$	$0.364 \pm$	$0.367 \pm$	$0.366 \pm$
	0.453	0.444	0.454	0.149	0.179	0.165

In conclusion, the $700 \times 700px$ heatmap model demonstrated better performance and higher consistency compared to the $256 \times 256px$ scanpath model. When compared to the base study [33], the model performance was similar ($700 \times 700px$ heatmap: Sensitivity-68%, Specificity-76%; Base model: Sensitivity-70%, Specificity-73%).

3.3 Image Content

In the three image content models, Table 3: Heatmap + RGB Image, Heatmap + GS Image, and RGB Image \times Heatmap, the performance declined, despite the performance of the model using only the $700 \times 700px$ heatmap. This decline may be due to a suboptimal architecture for these inputs or overfitting.

This result may have also been partially affected by the random image selection, resulting in an uneven distribution, with only 365 images shown more than 10 times and over 100 shown just twice (Fig. 4). This variability may have confused the model, though generalized eye movement patterns could still be beneficial, albeit requiring a larger dataset. As such, we suggest that conducting experiments with a more controlled image set could improve MCI prediction outcomes.

Table 3: Mean Sensitivity and specificity, in percentage, with respective standard deviation, for the models using image content as input, after the 10-fold nested cross-validation. The results from the $700 \times 700px$ Heatmap model are also included, for comparison.

	Heatmap (700×700)	Heatmap + RGB Image (700×700)	Heatmap + GS Image (700×700)	RGB Image \times Heatmap (700×700)
Sensitivity	68.42 ± 27.26	46.34 ± 47.80	48.78 ± 45.64	63.41 ± 33.78
Specificity	76.47 ± 27.64	62.16 ± 44.31	56.76 ± 44.87	43.24 ± 40.11

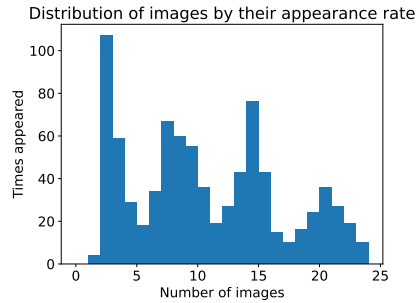


Fig. 4: Histogram of the distribution of the images used during the experiment.

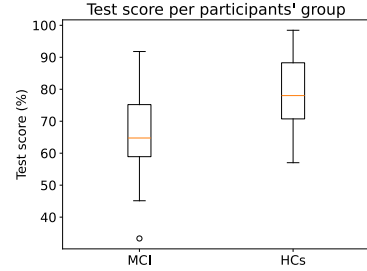


Fig. 5: Boxplot of the Test score for each of the experiment runs, separated by groups.

3.4 Trial and Test Score

We then analysed the two models related to task performance, Table 4: Heatmap + STS and Heatmap + ATS. Both models underperformed compared to the $700 \times 700px$ gaze heatmap model, though the trial score model's sensitivity was similar to the heatmap model, with reduced specificity. This suggests the model over-classifies HCs as MCI patients, possibly due to misclassifications in trial recognition. The entropy analysis of heatmaps (Table 2) also supports that incorrect trials exhibit smaller group differences, making trial scores less beneficial.

For the test score model, despite marked differences in average test scores between HCs (78.8%) and MCI patients (66.8%), there was significant overlap in performance, with MCI patients showing more variability (Fig. 5). Two factors limiting performance were the overlap in test scores and the small sample size of

Table 4: Mean Sensitivity and specificity, in percentage, with respective standard deviation, for the models using the scores from both each trial and the average of the whole test as input, after the 10-fold nested cross-validation. The results from the $700 \times 700px$ Heatmap model are also included, for comparison.

	Heatmap ($700 \times 700px$)	Heatmap + STS ($700 \times 700px$)	Heatmap + ATS ($700 \times 700px$)
Sensitivity	68.42 ± 27.26	68.29 ± 42.57	56.10 ± 49.72
Specificity	76.47 ± 27.64	45.95 ± 40.10	40.54 ± 51.64

around 35 participants per training set, which constrained the model’s learning capability.

4 Conclusions

This paper explored the use of deep learning to predict MCI using eye-tracking data and image content in a visual memory task. The $700 \times 700px$ heatmap model proved to be the most effective, achieving 68% sensitivity and 76% specificity. These results are comparable to those of the base study [33] (70% Sensitivity and 73% specificity), despite our study operating under more challenging conditions. However, the performance remains below diagnostic thresholds.

The models using image content performed significantly lower than expected. We believe this was primarily due to the large number of images relative to the sample size, which hindered the models’ ability to extract meaningful information.

Furthermore, incorporating trial-by-trial and overall task scores did not improve the results. This suggests that the gaze heatmaps may already encapsulate relevant information regarding trial-wise performance, rendering the direct use of task scores redundant.

Future work should prioritize standardizing the image set across all participants and increasing the sample size. Additionally, adapting the model for use with digital cameras could enhance accessibility for MCI screening.

Disclosure of Interests. All authors have no conflicts of interest.

References

1. Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C.K., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M.Y., Suk, H., Zhang, S., Suo, M., Tillet, P., Zhao, X., Wang, E., Zhou, K., Zou, R., Wang, X., Mathews, A., Wen, W., Chanan, G., Wu, P., Chintala, S.: Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2. p. 929–947. Association for Computing Machinery (2024). <https://doi.org/10.1145/3620665.3640366>
2. Bernstein Sideman, A., Al-Rousan, T., Tsoy, E., Piña Escudero, S.D., Pintado-Caipa, M., Kanjanapong, S., Mbakile-Mahlanza, L., Okada de Oliveira, M., De la Cruz-Puebla, M., Zygouris, S., et al.: Facilitators and barriers to dementia assessment and diagnosis: Perspectives from dementia experts within a global health context. *Frontiers in neurology* **13**, 769360 (2022). <https://doi.org/10.3389/fneur.2022.769360>

3. Bhargava, Y., Shetty, S.K., Baths, V.: Subjective cognitive decline prediction on imbalanced data using data-resampling and cost-sensitive training methods. *Procedia Computer Science* **235**, 1964–1979 (2024). <https://doi.org/10.1016/j.procs.2024.04.186>, international Conference on Machine Learning and Data Engineering (ICMLDE 2023)
4. Breitner, J.C.S.: Dementia—epidemiological considerations, nomenclature, and a tacit consensus definition. *Journal of Geriatric Psychiatry and Neurology* **19**(3), 129–136 (2006). <https://doi.org/10.1177/0891988706291081>, PMID: 16880354
5. Chen, Y., Power, M.C., Grodstein, F., Capuano, A.W., Lange-Maia, B.S., Moghtaderi, A., Stapp, E.K., Bhattacharyya, J., Shah, R.C., Barnes, L.L., et al.: Correlates of missed or late versus timely diagnosis of dementia in healthcare settings. *Alzheimer's & Dementia* **20**(8), 5551–5560 (2024). <https://doi.org/10.1002/alz.14067>
6. Coco, M.I., Merendino, G., Zappalà, G., Della Sala, S.: Semantic interference mechanisms on long-term visual memory and their eye-movement signatures in mild cognitive impairment. *Neuropsychology* **35**(5), 498 (2021). <https://doi.org/10.1037/neu0000734>
7. Dalmaijer, E.S., Mathôt, S., Van der Stigchel, S.: Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior research methods* **46**, 913–921 (2014). <https://doi.org/10.3758/s13428-013-0422-2>
8. Damiano, C., Walther, D.B.: Distinct roles of eye movements during memory encoding and retrieval. *Cognition* **184**, 119–129 (2019). <https://doi.org/10.1016/j.cognition.2018.12.014>
9. Frank, S.J., Frank, A.M.: Salient slices: Improved neural network training and performance with image entropy. *Neural Computation* **32**(6), 1222–1237 (2020). <https://doi.org/10.48550/arXiv.1907.12436>
10. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al.: Mild cognitive impairment. *The lancet* **367**(9518), 1262–1270 (2006). [https://doi.org/10.1016/S0140-6736\(06\)68542-5](https://doi.org/10.1016/S0140-6736(06)68542-5)
11. Hessels, R.S., Hooge, I.T.: Eye tracking in developmental cognitive neuroscience – the good, the bad and the ugly. *Developmental Cognitive Neuroscience* **40**, 100710 (2019). <https://doi.org/10.1016/j.dcn.2019.100710>
12. Jack Jr., C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Contributors, Elliott, C., Masliah, E., Ryan, L., Silverberg, N.: NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia* **14**(4), 535–562 (2018). <https://doi.org/10.1016/j.jalz.2018.02.018>
13. Krstić, K., Šošković, A., Ković, V., Holmqvist, K.: All good readers are the same, but every low-skilled reader is different: an eye-tracking study using pisa data. *European Journal of Psychology of Education* **33**, 521–541 (2018). <https://doi.org/10.1007/s10212-018-0382-0>
14. Li, A., Li, J., Zhang, D., Wu, W., Zhao, J., Qiang, Y.: Synergy through integration of digital cognitive tests and wearable devices for mild cognitive impairment screening. *Frontiers in Human Neuroscience* **17**, 1183457 (2023). <https://doi.org/10.3389/fnhum.2023.1183457>

15. Livingston, G., Huntley, J., Liu, K.Y., Costafreda, S.G., Selbæk, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., et al.: Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. *The Lancet* **404**(10452), 572–628 (2024). [https://doi.org/10.1016/S0140-6736\(24\)01296-0](https://doi.org/10.1016/S0140-6736(24)01296-0)
16. Madan, S., Lentzen, M., Brandt, J., Rueckert, D., Hofmann-Apitius, M., Fröhlich, H.: Transformer models in biomedicine. *BMC Medical Informatics and Decision Making* **24**(1), 214 (2024). <https://doi.org/10.1186/s12911-024-02600-5>
17. Mathôt, S., Schreij, D., Theeuwes, J.: Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods* **44**, 314–324 (2012). <https://doi.org/10.3758/s13428-011-0168-7>
18. NVIDIA, Vingelmann, P., Fitzek, F.H.: Cuda, release: 12.1 (2024), <https://developer.nvidia.com/cuda-toolkit>
19. Oladunni, T., Tossou, S., Haile, Y., Kidane, A.: Covid-19 county level severity classification with imbalanced class: A nearmiss under-sampling approach. *medRxiv* pp. 2021–05 (2021). <https://doi.org/10.1101/2021.05.21.21257603>
20. Ooms, K., Krassanakis, V.: Measuring the spatial noise of a low-cost eye tracker to enhance fixation detection. *Journal of Imaging* **4**(8) (2018). <https://doi.org/10.3390/jimaging4080096>
21. Organization, W.H., et al.: Global status report on the public health response to dementia (2021), <https://digitalcommons.fiu.edu/srreports/health/health/65/>
22. Petersen, R.C.: Mild cognitive impairment. *CONTINUUM: lifelong Learning in Neurology* **22**(2), 404–418 (2016). <https://doi.org/10.1212/CON.00000000000000313>
23. Rahane, A.A., Subramanian, A.: Measures of complexity for large scale image datasets. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAHC). pp. 282–287. IEEE (2020). <https://doi.org/10.1109/ICAHC48513.2020.9065274>
24. Reisberg, D.: The Oxford handbook of cognitive psychology, chap. 5. Eye Movements, pp. 69–82. OUP USA (2013). <https://doi.org/10.1093/oxfordhb/9780195376746.001.0001>
25. Roberts, R., Knopman, D.S.: Classification and epidemiology of mci. *Clinics in geriatric medicine* **29**(4), 753–772 (2013). <https://doi.org/10.1016/j.cger.2013.07.003>
26. Rukundo, O.: Effects of image size on deep learning. *Electronics* **12**(4), 985 (2023). <https://doi.org/10.3390/electronics12040985>
27. Sabbagh, M.N., Boada, M., Borson, S., Chilukuri, M., Doraiswamy, P., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A., Possin, K., et al.: Rationale for early diagnosis of mild cognitive impairment (mci) supported by emerging digital technologies. *The journal of prevention of Alzheimer's disease* **7**, 158–164 (2020). <https://doi.org/10.14283/jpad.2020.19>
28. Sabottke, C.F., Spieler, B.M.: The effect of image resolution on deep learning in radiography. *Radiology: Artificial Intelligence* **2**(1), e190015 (2020). <https://doi.org/10.1148/ryai.2019190015>, PMID: 33937810
29. Semma, A., Lazrak, S., Hannad, Y., Boukhani, M., El Kettani, Y.: Writer identification: The effect of image resizing on cnn performance. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **46**, 501–507 (2021). <https://doi.org/10.5194/isprs-archives-XLVI-4-W5-2021-501-2021>
30. Shah, H.A., Khalil, S., Andberg, S., Koivisto, A.M., Bednarik, R.: Eye tracking based detection of mild cognitive impairment: A review. *Information Fusion* **122**, 103202 (2025). <https://doi.org/10.1016/j.inffus.2025.103202>

31. Sims, S.D., Conati, C.: A neural architecture for detecting user confusion in eye-tracking data. In: Proceedings of the 2020 International Conference on Multimodal Interaction. p. 15–23. ICMI '20, Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3382507.3418828>
32. Souto, D., Kerzel, D.: Visual selective attention and the control of tracking eye movements: a critical review. *Journal of Neurophysiology* **125**(5), 1552–1576 (2021). <https://doi.org/10.1152/jn.00145.2019>, PMID: 33730516
33. Sriram, H., Conati, C., Field, T.: Classification of alzheimer's disease with deep learning on eye-tracking data. In: Proceedings of the 25th International Conference on Multimodal Interaction. pp. 104–113 (2023). <https://doi.org/10.1145/3577190.3614149>
34. Tadokoro, K., Yamashita, T., Fukui, Y., Nomura, E., Ohta, Y., Ueno, S., Nishina, S., Tsunoda, K., Wakutani, Y., Takao, Y., Miyoshi, T., Higashi, Y., Osakada, Y., Sasaki, R., Matsumoto, N., Kawahara, Y., Omote, Y., Takemoto, M., Hishikawa, N., Morihara, R., Abe, K.: Early detection of cognitive decline in mild cognitive impairment and alzheimer's disease with a novel eye tracking test. *Journal of the Neurological Sciences* **427**, 117529 (2021). <https://doi.org/10.1016/j.jns.2021.117529>
35. Thum, C.: Measurement of the entropy of an image with application to image focusing. *Optica Acta: International Journal of Optics* **31**(2), 203–211 (1984). <https://doi.org/10.1080/713821475>
36. Wolf, A., Tripanpitak, K., Umeda, S., Otake-Matsuura, M.: Eye-tracking paradigms for the assessment of mild cognitive impairment: a systematic review. *Frontiers in Psychology* **14** (2023). <https://doi.org/10.3389/fpsyg.2023.1197567>