

Article

Not peer-reviewed version

---

# AI-based Facial Emotion Analysis for Early and Differential Diagnosis of Dementia

---

[Letizia Bergamasco](#)<sup>\*</sup>, [Anita Coletta](#), [Gabriella Olmo](#), [Aurora Cermelli](#), [Elisa Rubino](#), [Innocenzo Rainero](#)

Posted Date: 16 June 2025

doi: 10.20944/preprints202506.1295.v1

Keywords: artificial intelligence; machine learning; facial expression recognition; affective computing; mild cognitive impairment; dementia; Alzheimer's disease









Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

## Article

# AI-Based Facial Emotion Analysis for Early and Differential Diagnosis of Dementia

Letizia Bergamasco <sup>1,2,\*</sup> , Anita Coletta <sup>1</sup> , Gabriella Olmo <sup>1</sup> , Aurora Cermelli <sup>3,4</sup> ,  
Elisa Rubino <sup>3,4</sup>  and Innocenzo Rainero <sup>3,4</sup> 

<sup>1</sup> Department of Control and Computer Engineering, Politecnico di Torino, 10129 Turin, Italy

<sup>2</sup> LINKS Foundation, 10138 Turin, Italy

<sup>3</sup> Center for Alzheimer's Disease and Related Dementias, Department of Neuroscience and Mental Health, A.O.U. Città della Salute e della Scienza di Torino, 10126 Turin, Italy

<sup>4</sup> Department of Neuroscience "Rita Levi Montalcini", University of Torino, 10126 Turin, Italy

\* Correspondence: letizia.bergamasco@polito.it

**Abstract:** Early and differential diagnosis of dementia is essential for timely and targeted care. This study investigated the feasibility of using an artificial intelligence (AI)-based system to discriminate between different stages and etiologies of dementia by analyzing facial emotions. We collected video recordings of 64 participants exposed to standardized audio-visual stimuli. Facial emotion features in terms of valence and arousal were extracted and used to train machine learning models on multiple classification tasks, including distinguishing individuals with mild cognitive impairment (MCI) and overt dementia from healthy controls (HC), and differentiating Alzheimer's disease (AD) from other types of cognitive impairment. The system achieved a cross-validation accuracy of 76.0% for MCI vs HC, 73.6% for dementia vs HC, and 64.1% in the three-class classification (MCI vs dementia vs HC). Among cognitively impaired individuals, 75.4% accuracy was reached in distinguishing AD from other etiologies. These results demonstrated the potential of AI-driven facial emotion analysis as a non-invasive tool for early detection of cognitive impairment, and for supporting differential diagnosis of AD in clinical settings.

**Keywords:** artificial intelligence; machine learning; facial expression recognition; affective computing; mild cognitive impairment; dementia; Alzheimer's disease

## 1. Introduction

Dementia is a term used to describe a syndrome characterized by a progressive deterioration of cognitive functions and behavioral disturbances. It is recognized by the World Health Organization (WHO) as one of the main causes of disability and loss of autonomy in the elderly population at a global level [1]. Alzheimer's disease (AD) is the most common form of neurodegenerative dementia, typically presenting at onset with short-term memory impairment [2]. Less frequent types of dementia include vascular dementia (VD) [3], frontotemporal dementia (FTD) [4], dementia with Lewy bodies (DLB) [5], and mixed forms.

Early and differential diagnosis is essential for accessing dementia care and support, as well as for enrolling individuals in clinical trials. At present, therapeutic approaches are primarily aimed at providing symptomatic relief [6]. However, the Food and Drug Administration has recently approved disease-modifying therapies for AD [7,8], showing clinical efficacy only when administered during the earliest stages of the disease.

From a clinical perspective, Alzheimer's disease is a progressive continuum that begins with an asymptomatic phase, progresses to a stage of mild cognitive impairment (MCI), and ultimately culminates in the onset of overt dementia [9]. In the MCI stage, the first symptoms occur, but most patients can live independently and their daily activities are not heavily affected. This phase is thus considered an important time window for detecting and diagnosing cognitive decline at an early stage.

The diagnosis of MCI and dementia typically relies on a combination of medical history, neuropsychological assessments, neuroimaging, and laboratory tests, which are often costly, invasive, and require specialized clinical expertise. Therefore, there is a pressing need for the development of accessible, cost-effective approaches for the early detection of cognitive impairment (CI). In this context, facial expressions may carry important diagnostic information; altered facial expressivity is commonly observed in individuals with CI, and these changes tend to differ depending on the type and stage of dementia [10,11]. This makes the assessment of facial emotional expression potentially useful for distinguishing between different forms of dementia.

Computer vision approaches, especially deep learning (DL)-based ones, have shown promising results in analyzing facial expressions to support an early and accurate diagnosis of CI. Some studies fed facial images into DL models, in an end-to-end fashion. Sun et al. [12] achieved 90.63% accuracy in distinguishing MCI from healthy controls (HC) with a Multi-branch Classifier-Video Vision Transformer; they used a subset (83 MCI and 64 HC) of the I-CONNECT dataset [13], containing semi-structured interviews. Umeda-Kameyama et al. [14] reached 92.56% accuracy in CI detection with an Xception DL model; their dataset encompassed 484 face images (including 121 patients with AD and 117 HC).

Other studies extracted facial-related features to perform CI detection. Zheng et al. [15] reached a dementia detection accuracy of 79% with Histogram of Oriented Gradients features; 71% with Action Units (AUs); and 66% with face mesh features. They used a subset of the PROMPT dataset [16], including 447 videos of 117 subjects (HC and dementia patients). Particularly of interest are facial features representing emotions, that not only have shown good performance, but also facilitate model interpretability for clinicians, since emotional regulation is directly affected by dementia. Fei et al. [17] extracted categorical emotion features with a DL-based model, and used them to train a Support Vector Machine (SVM), reaching 73.3% accuracy in distinguishing CI and HC subjects. For their dataset, the authors enrolled 61 elderly people (36 CI and 25 HC), and developed an interface to record facial videos while displaying video emotional stimuli. In a previous study of our research group [18], we proposed a novel approach for CI detection; while our aim was similar to that of [17], we integrated a dimensional model of affect for a more comprehensive emotion representation, and used standardized emotion elicitation stimuli for data collection. We achieved 76.7% accuracy in CI detection on facial videos recorded from 60 subjects (32 CI and 28 HC).

The above mentioned studies focused specifically on MCI or dementia detection, or grouped together MCI and dementia subjects to perform CI detection. However, very few works in literature targeted the joint differentiation of different stages of the disease (e.g., MCI and overt dementia), or different underlying etiologies (e.g., AD and other forms of neurodegenerative condition). A recent work by Okunishi et al. [19] proposed a methodology to detect MCI and dementia based on AUs, eight emotion categories, valence-arousal, and face embeddings. By extracting and combining all these features from video recordings, they achieved 86.2% accuracy on dementia detection and 83.4% accuracy on MCI detection on a selected subset of the PROMPT dataset [16]. Chu et al. [20] recruited 95 participants (MCI: 41, mild to moderate dementia: 54) and recorded them during the Short Portable Mental Status Questionnaire process. They performed binary classification of MCI and dementia with DL models trained on visual and speech features, reaching 76.0% accuracy (88% when excluding depression and anxiety). On the other hand, Jiang et al. [21] conducted a comprehensive study on 493 individuals (including HC, MCI due to AD, MCI due to other etiologies, dementia due to AD, dementia due to other etiologies, subjective CI) video-recorded during a passive memory test. Through facial emotion analysis via DL, the authors were able to differentiate CI participants from HC, but not to differentiate the underlying etiologies.

Based on our previous experience [18], the aim of this work was to investigate the automatic detection of both MCI and overt dementia conditions using facial emotion features extracted from video recordings. Moreover, we also aimed to use our proposed system to automatically discriminate AD from other forms of cognitive impairment. To the best of our knowledge, this is the first study

to propose an automated method to differentiate between diverse etiologies of dementia based on facial emotion analysis from video data. For our analysis, we collected video data from subjects whose diagnosis was supported by relevant biomarkers, including AD biomarkers in the cerebrospinal fluid. Our results showed that good detection performance can be reached not only for dementia patients, but also for MCI patients, and in the discrimination of AD, showing promise for the support to early and differential diagnosis of dementia.

The remainder of the paper is structured as follows. Section 2 describes the collected dataset, the architecture of the employed system for CI detection, and the performed experiments; Section 3 presents the experimental results; finally, Section 4 discusses the results and outlines future research directions.

## 2. Materials and Methods

### 2.1. Collected Data

Our dataset encompassed video recordings from subjects exposed to an emotion elicitation video. The full protocol used for data collection was introduced in [18]. In detail, the emotion elicitation video was created using images and sounds from two databases widely used in affective stimulation research, i.e., IAPS (International Affective Picture System [22]) and IADS-2 (International Affective Digitized Sounds-2 [23]) respectively. This ensured that the employed emotional stimuli were standardized. Specifically, 28 pairs made of an image and a sound with similar valence and arousal were displayed in a fixed, randomly selected sequence.

According to the protocol, the subject was positioned in front of a laptop, which simultaneously showed the emotion elicitation video and recorded the subject's facial expression via an external USB webcam (Logitech C920, with 1080p resolution and 30 fps frame rate). A nearby external Bluetooth speaker was used to ensure high-quality sound output. The experiment was set up using PsychoPy v2022.2.4 software [24], that enabled the synchronized presentation of emotional stimuli and simultaneous webcam recording.

The emotion elicitation video lasts about 8 minutes. It starts with a webcam calibration phase of 10 s, and a welcome title displayed for 5.5 s. Then, there is the sequence of 28 audio-visual stimuli: each trial begins with a 10-second countdown, followed by a 1-second display of a central cross; subsequently, the image is presented for 6 seconds while the associated sound plays simultaneously. Once the audiovisual stimuli conclude, the recording stops and a concluding title is displayed for 1 second.

Starting from the data presented in [18], we collected additional video recordings to expand the database. In this way, we obtained data from a total of 64 participants, among which 28 HC subjects, 26 diagnosed with MCI (13: due to AD; 13: other types), and 10 diagnosed with overt dementia (4: AD; 6: other types). Diagnoses of AD were based on the NIA-AA (National Institute on Aging and the Alzheimer's Association) AT(N) criteria, incorporating biomarkers for amyloid (A), tau (T), and neurodegeneration (N) [25–27]. Non-AD forms of CI included different etiologies such as FTD, DLB, VD, mixed or not specified forms; individuals with subjective CI were also included. Diagnoses of FTD, DLB, VD and other neurocognitive disorders were made based on established diagnostic criteria specific to each condition [28–31]. The experiments were conducted in a designated room at Molinette Hospital – A.O.U. Città della Salute e della Scienza di Torino. Table 1 provides a summary of the demographic characteristics and key clinical information.



Table 1. Participants’ demographics and key clinical information.

	MCI	Overt dementia	Healthy controls
Number of subjects	26	10	28
Age (mean ±standard deviation)	68.2 ±9.3	72.9 ±3.8	58.8 ±6.9
Sex (number of females, %)	10 (38.5%)	6 (60.0%)	14 (50.0%)
Ethnicity	Caucasian	Caucasian	Caucasian
Years of education (mean ±standard deviation)	13.7 ±4.6	10.4 ±5.4	15.6 ±4.8
MMSE score (mean ±standard deviation)	25.8 ±3.6	18.8 ±5.5	29.2 ±1.2
MoCA score (mean ±standard deviation)	20.0 ±4.4	14.0 ±3.6	25.4 ±2.2
Differential CI diagnosis	13: due to AD; 13: other types	4: AD; 6: other types	No cognitive impairment

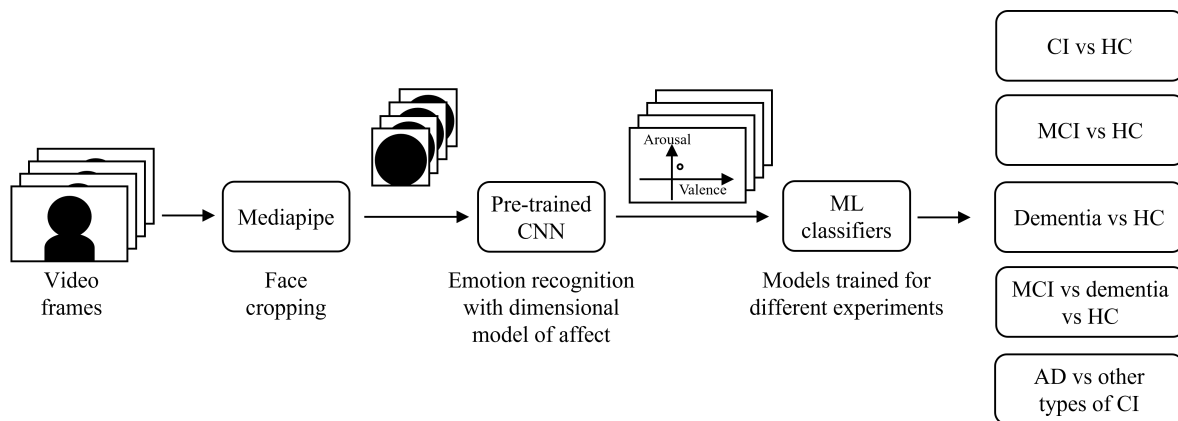
Participants with CI were selected from individuals seeking diagnosis of cognitive disorders at the Center for Alzheimer’s Disease and Related Dementias at the Department of Neuroscience and Mental Health, A.O.U. Città della Salute e della Scienza University Hospital (Turin, Italy). The diagnoses and differential diagnoses of CI were performed based on a comprehensive neurological and neurocognitive evaluation, including neuropsychological testing, brain imaging (MRI and 18F-fluorodeoxyglucose PET), and lumbar puncture for cerebrospinal fluid biomarker analysis ( $A\beta 42$ ,  $A\beta 42/A\beta 40$ , total tau, and phosphorylated tau 181).

The classification between MCI and overt dementia subjects was based on the cognitive evaluations conducted by a trained neuropsychologist (A.Ce.), using the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), and assessments of functional independence through the Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) scales. Participants with CI were classified as MCI-affected if they met the criteria of  $MMSE \geq 20$ ,  $ADL = 6/6$  and  $IADL \geq 6/8$ . Those with  $MMSE < 20$  or  $ADL < 6/6$  or  $IADL < 6/8$  were categorized as affected by overt dementia.

On the other hand, HC subjects consisted of volunteers aged between 40 and 80 years. Exclusion criteria included the presence of neurological or psychiatric disorders, as well as any other condition that could hinder participation in the experiment (e.g., blindness). HC subjects also underwent neuropsychological assessment, since eligibility required  $MMSE \geq 26/30$ ,  $ADL = 6/6$  and  $IADL = 8/8$ .

2.2. System Architecture and Data Processing

For our experiments, we employed an architecture based on the system presented in [18] and adapted to perform different CI detection tasks. It is characterized by two subsequent parts: (i) firstly, it obtains the evolution of the emotions from the collected videos, in terms of valence and arousal; (ii) secondly, it uses the emotions data to train a machine learning (ML) model for the selected CI detection task. A scheme of the developed system is illustrated in Figure 1.



**Figure 1.** Scheme of the system architecture.

For each video we extracted all frames ( $\sim 14k$ ), and we cropped the subject's face with MediaPipe [32] Holistic solution, with a  $224 \times 224$  pixels size. Each frame was processed by two pre-trained Convolutional Neural Networks (CNNs) to perform facial emotion recognition, and estimate valence and arousal values, respectively. Emotions were therefore represented with a dimensional model, in particular the circumplex model [33], using a circular space defined by two affect dimensions: valence, indicating if an emotion is positive or negative, and arousal, indicating its intensity. With respect to categorical models of emotions (e.g., Ekman's Basic Emotions model [34] with six basic emotions), dimensional models can capture all possible emotion nuances.

The CNN models, introduced and detailed in [18], were trained on the AffectNet dataset [35], showing a performance on valence and arousal prediction comparable to the AffectNet benchmark [35]. For this reason, also in this work, these CNN models were deemed adequate to be used within our system for CI detection. The exploration of more complex DL models for facial emotion recognition was beyond the scope of this paper and is left to future work.

The resulting valence and arousal series were concatenated into a single feature vector, representing how participants' emotional states changed over the course of the experiment. These feature vectors were used to train different classification algorithms (detailed in Section 2.4), according to the different experiments, as discussed in the following Section.

### 2.3. Experiments

With the same system pipeline outlined in Section 2.2, we performed five different experiments.

1. *CI vs HC*. In this experiment, all CI subjects were grouped together and compared with the HC group by performing a binary classification. This allowed to validate the generalization capability of the proposed algorithm when tested on this enlarged dataset against that in [18]. The considered dataset included 64 subjects: 36 CI (26 MCI + 10 overt dementia) and 28 HC.
2. *MCI vs HC*. Among the CI subjects, in this experiment we selected only those clinically diagnosed as MCI. The objective was to investigate if differences with respect to HC could be spotted also in the earlier phase of the disease. The considered dataset included 54 subjects: 26 MCI and 28 HC. We performed binary classification to distinguish these two classes.
3. *Dementia vs HC*. In contrast to the previous experiment, in this one we selected only overt dementia patients, to focus on the differences with respect to HC that could be spotted at a later phase of the disease. The considered dataset included 38 subjects: 10 overt dementia and 28 HC. We performed binary classification to distinguish these two classes.
4. *MCI vs dementia vs HC*. In this experiment, we compared the three different classes of subjects, according to the level of severity of the disease. The considered dataset included 64 subjects: 26 MCI, 10 overt dementia, and 28 HC. We moved from a binary to a multiclass classification problem to distinguish these three classes. It must be noticed that the dataset was imbalanced

across classes, as the overt dementia class included fewer subjects compared to the other two groups.

5. *AD vs other types of CI.* In this last experiment, the aim was to investigate differences in facial emotion responses among individuals with different types of CI. Specifically, we grouped together patients diagnosed with AD, and compared them to the broader group of individuals with other forms of CI. This approach was motivated by the fact that AD is the most common cause of dementia, and a differential diagnosis distinguishing AD from other etiologies is of critical clinical importance. The considered dataset included 36 subjects: 26 MCI (13: due to AD; 13: other types), and 10 overt dementia (4: AD, 6: other types). We considered two classes: AD (17 subjects), and other types of CI (19 subjects). We performed binary classification to distinguish these two classes.

#### 2.4. Model Selection and Evaluation

Machine Learning classifiers were implemented using the scikit-learn Python library [36]. For binary classification tasks (experiments 1,2,3,5), considering also the limited size of our dataset, we selected K-Nearest Neighbors (KNN), Logistic Regression (LR), and Support Vector Machine (SVM) algorithms. KNN was optimized via grid search over the number of neighbors (3, 5, 7) and choice of distance metric (Euclidean, Manhattan, and Chebyshev). For LR, we applied an L2 regularization term with the “liblinear” solver, a tolerance for stopping criteria of  $10^{-4}$ , and tuned the inverse of regularization strength parameter  $C$  across a range of powers of 10 from  $10^{-4}$  to  $10^4$ . SVM was configured with a linear kernel, a tolerance of  $10^{-3}$ , and was similarly tuned on the regularization parameter  $C$ , again spanning powers of 10 from  $10^{-4}$  to  $10^4$ .

For multiclass classification tasks (experiment 4), the ML models were implemented similarly, with a few adjustments to accommodate the multiclass setting. The scikit-learn KNN estimator supports multiclass problems, therefore no modifications were required. For the LR estimator, we changed the “liblinear” solver to “lbfgs”, which is able to handle multinomial loss. For the SVM estimator, multiclass classification is managed using a one-vs-one strategy, so we employed it following the same procedure as for the binary classification cases.

In order to provide an unbiased estimate of the ML models’ generalization error, we used the nested cross-validation (NCV) technique. In fact, with datasets of limited size, standard cross-validation (CV) used for both hyperparameter tuning and performance estimation may produce over-optimistic results. Instead, NCV uses an inner CV loop to search for the best set of model parameters, and an outer CV loop to evaluate the final model performance independently. This separation leads to a less biased and more realistic estimate of the model’s generalization error. In all our experiments, we performed a 5-fold CV in the outer loop, and a 3-fold CV in the inner loop. To maintain class distribution consistency across folds, both the outer and inner CV loops used stratified k-fold cross-validation as implemented in scikit-learn, ensuring that each fold contained approximately the same class proportions as the original dataset.

This procedure was applied to all the ML models under consideration. The model achieving the highest average accuracy across the NCV outer folds was selected as the best-performing one. For this selected model, the optimal set of hyperparameters was determined as the combination most frequently chosen across the outer folds.

### 3. Results

The results provided by the different classification experiments involving CI and HC subjects (experiments 1,2,3,4) are shown in Table 2. In summary, when classifying all CI subjects vs HC subjects, the best-performing model was a KNN with 73.6% accuracy. Instead, when considering different stages of CI separately, another KNN model reached the best accuracy of 76.0% in classifying MCI vs HC subjects; on the other hand, the best accuracy of 73.6% on the classification of dementia and HC subjects was reached by an SVM model. Lastly, for the multiclass problem MCI vs dementia vs HC, a cross-validation accuracy of 64.1% was reached by a KNN model.

**Table 2.** Cross-validation accuracy for different classification experiments involving CI and HC subjects (mean ±standard deviation)

Experiment	Model	Parameters	Accuracy
CI vs HC	KNN	3 neighbors, Manhattan distance	<b>0.736 ±0.102</b>
	LR	L2 penalty, tolerance=0.0001, C=0.001	0.623 ±0.139
	SVM	linear kernel, tolerance=0.001, C=0.01	0.624 ±0.092
MCI vs HC	KNN	3 neighbors, Manhattan distance	<b>0.760 ±0.041</b>
	LR	L2 penalty, tolerance=0.0001, C=0.001	0.684 ±0.114
	SVM	linear kernel, tolerance=0.001, C=0.001	0.667 ±0.069
Dementia vs HC	KNN	3 neighbors, Euclidean distance	0.732 ±0.097
	LR	L2 penalty, tolerance=0.0001, C=0.1	0.654 ±0.145
	SVM	linear kernel, tolerance=0.001, C=0.0001	<b>0.736 ±0.018</b>
MCI vs dementia vs HC	KNN	5 neighbors, Manhattan distance	<b>0.641 ±0.103</b>
	LR	L2 penalty, tolerance=0.0001, C=0.01	0.591 ±0.104
	SVM	linear kernel, tolerance=0.001, C=0.1	0.578 ±0.077

Results about the classification between AD and other types of CI (experiment 5) are shown in Table 3. It can be noticed that the best accuracy of 75.4% was reached by a KNN model.

**Table 3.** Cross-validation accuracy for the classification of AD versus other types of CI (mean ±standard deviation)

Experiment	Model	Parameters	Accuracy
AD vs other types of CI	KNN	5 neighbors, Chebyshev distance	<b>0.754 ±0.128</b>
	LR	L2 penalty, tolerance=0.0001, C=0.0001	0.586 ±0.171
	SVM	linear kernel, tolerance=0.001, C=0.01	0.643 ±0.090

4. Discussion

The results of our study suggest that our models are capable of distinguishing CI and HC subjects using exclusively the facial emotion data collected with our protocol (Table 2). With respect to [18] (reaching 76.7% accuracy for the classification of CI vs HC subjects), the current study was based on an



enlarged dataset and achieved a still good accuracy of 73.6%, suggesting that the method maintained good performance despite the increased heterogeneity of the CI population. Nonetheless, further data collection will allow us to more extensively assess the generalizability of these findings.

The obtained results in CI detection indicate that emotional features extracted from our collected videos can be an effective tool to support the screening of CI. This further highlights the potential of investigating the emotional dimension in the context of early diagnosis; this aspect is currently underexplored in neuropsychological tests such as MMSE and MoCA, but was proved to be informative also in other literature studies [17–19,21].

When analyzing classification performance across different stages of cognitive decline, i.e., when considering separately the cases of MCI vs HC, and dementia vs HC, the accuracy was comparable or even increased with respect to the previous experiment: 76.0% and 73.6%, respectively. The observed difference in performance between the cases of MCI vs HC (76.0%) and dementia vs HC (73.6%) should be interpreted with caution, as a more reliable comparison would require the two classes to have comparable sample sizes; anyway, our results show that the accuracy is still satisfactory when considering only MCI patients (without the ones with overt dementia). Therefore, our method results to be promising in supporting the early diagnosis of cognitive impairment, since it is effective not only with overt dementia, but also in subjects in an earlier stage of CI.

In the multiclass classification of MCI vs dementia vs HC, the accuracy decreased compared to the binary classification cases (MCI vs HC, and dementia vs HC), yet it remained relatively high (64.1%) considering the increased complexity of the task. This is not surprising, as distinguishing CI from HC subjects is generally easier than the more fine-grained differentiation between different stages of the dementia continuum. Interestingly, this limitation was observed also in studies applying ML to neuroimaging data for the diagnosis and prognosis of CI and dementia. Indeed, Pellegrini et al. [37] reported that, although ML methods show acceptable accuracy in distinguishing overt AD from HC, their performance drops when trying to differentiate MCI from AD, MCI from HC, or for predicting MCI conversion to AD.

One of the most interesting findings of this study is related to the discrimination of AD vs other types of CI, reaching a cross-validation accuracy of 75.4% (Table 3). This result is very promising, as it is likely possible to perform a differential diagnosis of AD in a non-invasive way, by exploiting facial emotion analysis.

A comprehensive performance comparison with related studies is currently challenging, due to differences in the video datasets used, the corresponding data collection protocols, and the varying definitions of the ML classification tasks. A distinguishing feature of the present work, in contrast to others such as [17,21], lies in the use of an emotion elicitation protocol grounded in standardized, extensively validated stimuli [22,23], enhancing its facility of adoption and generalizability. Moreover, our study adopted a dimensional model of emotions, which provides a more comprehensive way of characterizing affective states compared to the categorical approach used by Fei et al. [17], who also focused on CI detection based solely on facial emotions. Most importantly, among the strengths of the study is the availability of well-characterized ground truth classification for CI subjects. Unlike other studies, in which the type of CI of the subjects was often assumed and not properly confirmed [14,20], in our study the diagnosis process was based on a comprehensive evaluation, as explained in Section 2.1, and AD diagnoses were supported by biomarkers from cerebrospinal fluid.

While the experimental results were promising, this study presents some limitations. First of all, the collected dataset had a limited sample size (64 participants), all of Caucasian ethnicity and recruited from a single clinical center within a specific setting. To validate the generalization capability of the presented results, in the future we aim to collect data from more subjects, possibly including diverse ethnicities, and in a multicenter perspective. Secondly, the dataset was not balanced in all the classes, since the dementia class included less subjects than the MCI and the HC classes. This imbalance resulted from the recruitment process: patients were enrolled based on eligibility criteria during their visits to our Center for Alzheimer's Disease and Related Dementias, without prior knowledge of

their clinical diagnosis. More experiments could be performed in the future, when more data will be collected, and more advanced techniques to deal with class imbalance could be explored.

In addition, the overall system performance was largely influenced by the method used to extract facial emotion features, which still offers room for performance improvement and could contribute to an even more precise encoding of emotional information. Therefore, future research directions include the exploration of more complex DL architectures that demonstrated good performance in facial emotion recognition tasks, such as attention mechanisms [38,39]. Moreover, we would like to investigate the integration of more facial features beyond valence and arousal, as the combination of multiple feature types was shown to be effective in previous studies [19].

**Author Contributions:** L.B.: writing—original draft preparation, data curation, formal analysis, investigation, methodology, visualization, validation. A.Co. (Anita Coletta): formal analysis, investigation, software, writing—review and editing. G.O.: conceptualization, funding acquisition, project administration, supervision, writing—review and editing. A.Ce. (Aurora Cermelli): data curation, investigation, writing—review and editing. E.R.: conceptualization, funding acquisition, project administration, writing—review and editing. I.R.: conceptualization, funding acquisition, supervision, resources, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Fondazione CRT grant number 105128 / 2023.0366.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of A.O.U. Città della Salute e della Scienza di Torino (approval number 0114576).

**Informed Consent Statement:** Written informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The dataset presented in this article is not readily available due to privacy restrictions.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. World Health Organization. *Global status report on the public health response to dementia*; 2021.
2. Scheltens, P.; Blennow, K.; Breteler, M.M.; De Strooper, B.; Frisoni, G.B.; Salloway, S.; Van der Flier, W.M. Alzheimer's disease. *The Lancet* **2016**, *388*, 505–517. [https://doi.org/10.1016/S0140-6736\(15\)01124-1](https://doi.org/10.1016/S0140-6736(15)01124-1).
3. T O'Brien, J.; Thomas, A. Vascular dementia. *The Lancet* **2015**, *386*, 1698–1706. [https://doi.org/10.1016/S0140-6736\(15\)00463-8](https://doi.org/10.1016/S0140-6736(15)00463-8).
4. Bang, J.; Spina, S.; Miller, B.L. Frontotemporal dementia. *The Lancet* **2015**, *386*, 1672–1682. [https://doi.org/10.1016/S0140-6736\(15\)00461-4](https://doi.org/10.1016/S0140-6736(15)00461-4).
5. Walker, Z.; Possin, K.L.; Boeve, B.F.; Aarsland, D. Lewy body dementias. *The Lancet* **2015**, *386*, 1683–1697. [https://doi.org/10.1016/S0140-6736\(15\)00462-6](https://doi.org/10.1016/S0140-6736(15)00462-6).
6. Koyama, A.; Okereke, O.I.; Yang, T.; Blacker, D.; Selkoe, D.J.; Grodstein, F. Plasma amyloid- $\beta$  as a predictor of dementia and cognitive decline: a systematic review and meta-analysis. *Archives of neurology* **2012**, *69*, 824–831. <https://doi.org/10.1001/archneurol.2011.1841>.
7. Alexander, G.C.; Emerson, S.; Kesselheim, A.S. Evaluation of aducanumab for Alzheimer disease: scientific evidence and regulatory review involving efficacy, safety, and futility. *Jama* **2021**, *325*, 1717–1718. <https://doi.org/10.1001/jama.2021.3854>.
8. van Dyck, C.H.; Swanson, C.J.; Aisen, P.; Bateman, R.J.; Chen, C.; Gee, M.; Kanekiyo, M.; Li, D.; Reyderman, L.; Cohen, S.; et al. Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **2023**, *388*, 9–21. <https://doi.org/10.1056/NEJMoa2212948>.
9. Albert, M.S.; DeKosky, S.T.; Dickson, D.; Dubois, B.; Feldman, H.H.; Fox, N.C.; Gamst, A.; Holtzman, D.M.; Jagust, W.J.; Petersen, R.C.; et al. The diagnosis of mild cognitive impairment 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 & dementia* **2011**, 7, 270–279. <https://doi.org/10.1016/j.jalz.2011.03.008>.
10. Chen, K.H.; Lwi, S.J.; Hua, A.Y.; Haase, C.M.; Miller, B.L.; Levenson, R.W. Increased subjective experience of non-target emotions in patients with frontotemporal dementia and Alzheimer's disease. *Current Opinion in Behavioral Sciences* **2017**, 15, 77–84. <https://doi.org/10.1016/j.cobeha.2017.05.017>.
  11. Pressman, P.S.; Chen, K.H.; Casey, J.; Sillau, S.; Chial, H.J.; Filley, C.M.; Miller, B.L.; Levenson, R.W. Incongruences between facial expression and self-reported emotional reactivity in frontotemporal dementia and related disorders. *The Journal of neuropsychiatry and clinical neurosciences* **2023**, 35, 192–201. <https://doi.org/10.1176/appi.neuropsych.21070186>.
  12. Sun, J.; Dodge, H.H.; Mahoor, M.H. MC-ViT: Multi-branch Classifier-ViT to detect Mild Cognitive Impairment in older adults using facial videos. *Expert Systems with Applications* **2024**, 238, 121929. <https://doi.org/10.1016/j.eswa.2023.121929>.
  13. Dodge, H.H.; Yu, K.; Wu, C.Y.; Pruitt, P.J.; Asgari, M.; Kaye, J.A.; Hampstead, B.M.; Struble, L.; Potempa, K.; Lichtenberg, P.; et al. Internet-Based Conversational Engagement Randomized Controlled Clinical Trial (I-CONNECT) Among Socially Isolated Adults 75+ Years Old With Normal Cognition or Mild Cognitive Impairment: Topline Results. *The Gerontologist* **2023**, 64, gnad147. <https://doi.org/10.1093/geront/gnad147>.
  14. Umeda-Kameyama, Y.; Kameyama, M.; Tanaka, T.; Son, B.K.; Kojima, T.; Fukasawa, M.; Iizuka, T.; Ogawa, S.; Iijima, K.; Akishita, M. Screening of Alzheimer's disease by facial complexion using artificial intelligence. *Aging (Albany NY)* **2021**, 13, 1765–1772. <https://doi.org/10.18632/aging.202545>.
  15. Zheng, C.; Bouazizi, M.; Ohtsuki, T.; Kitazawa, M.; Horigome, T.; Kishimoto, T. Detecting Dementia from Face-Related Features with Automated Computational Methods. *Bioengineering* **2023**, 10. <https://doi.org/10.3390/bioengineering10070862>.
  16. Kishimoto, T.; Takamiya, A.; Liang, K.; Funaki, K.; Fujita, T.; Kitazawa, M.; Yoshimura, M.; Tazawa, Y.; Horigome, T.; Eguchi, Y.; et al. The project for objective measures using computational psychiatry technology (PROMPT): Rationale, design, and methodology. *Contemporary Clinical Trials Communications* **2020**, 19, 100649. <https://doi.org/10.1016/j.conctc.2020.100649>.
  17. Fei, Z.; Yang, E.; Yu, L.; Li, X.; Zhou, H.; Zhou, W. A Novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly. *Neurocomputing* **2022**, 468, 306–316. <https://doi.org/10.1016/j.neucom.2021.10.038>.
  18. Bergamasco, L.; Lorenzo, F.; Coletta, A.; Olmo, G.; Cermelli, A.; Rubino, E.; Rainero, I. Automatic Detection of Cognitive Impairment through Facial Emotion Analysis. SSRN preprint, <https://doi.org/10.2139/ssrn.5192464>.
  19. Okunishi, T.; Zheng, C.; Bouazizi, M.; Ohtsuki, T.; Kitazawa, M.; Horigome, T.; Kishimoto, T. Dementia and MCI Detection Based on Comprehensive Facial Expression Analysis From Videos During Conversation. *IEEE Journal of Biomedical and Health Informatics* **2025**, 29, 3537–3548. <https://doi.org/10.1109/JBHI.2025.3526553>.
  20. Chu, C.S.; Wang, D.Y.; Liang, C.K.; Chou, M.Y.; Hsu, Y.H.; Wang, Y.C.; Liao, M.C.; Chu, W.T.; Lin, Y.T. Automated Video Analysis of Audio-Visual Approaches to Predict and Detect Mild Cognitive Impairment and Dementia in Older Adults. *Journal of Alzheimer's Disease* **2023**, 92, 875–886. <https://doi.org/10.3233/JAD-220999>.
  21. Jiang, Z.; Seyedi, S.; Haque, R.U.; Pongos, A.L.; Vickers, K.L.; Manzanares, C.M.; Lah, J.J.; Levey, A.I.; Clifford, G.D. Automated analysis of facial emotions in subjects with cognitive impairment. *Plos one* **2022**, 17, e0262527. <https://doi.org/10.1371/journal.pone.0262527>.
  22. Lang, P.J.; Bradley, M.M.; Cuthbert, B.N. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report Technical Report A-8, University of Florida, NIMH Center for the Study of Emotion and Attention, 2008.
  23. Bradley, M.M.; Lang, P.J. The International Affective Digitized Sounds (IADS-2): Affective Ratings of Sounds and Instruction Manual. Technical Report Technical Report B-3, University of Florida, NIMH Center for the Study of Emotion and Attention, 2007.
  24. Peirce, J.; Gray, J.R.; Simpson, S.; MacAskill, M.; Höchenberger, R.; Sogo, H.; Kastman, E.; Lindeløv, J.K. PsychoPy2: Experiments in behavior made easy. *Behavior research methods* **2019**, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>.
  25. 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.; et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & dementia* **2018**, 14, 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>.

26. Jack Jr, C.R.; Andrews, J.S.; Beach, T.G.; Buracchio, T.; Dunn, B.; Graf, A.; Hansson, O.; Ho, C.; Jagust, W.; McDade, E.; et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia* **2024**, *20*, 5143–5169. <https://doi.org/10.1002/alz.13859>.
27. Frisoni, G.B.; Festari, C.; Massa, F.; Ramusino, M.C.; Orini, S.; Aarsland, D.; Agosta, F.; Babiloni, C.; Borroni, B.; Cappa, S.F.; et al. European intersocietal recommendations for the biomarker-based diagnosis of neurocognitive disorders. *The Lancet Neurology* **2024**, *23*, 302–312. [https://doi.org/10.1016/S1474-4422\(23\)00447-7](https://doi.org/10.1016/S1474-4422(23)00447-7).
28. Rascovsky, K.; Hodges, J.R.; Knopman, D.; Mendez, M.F.; Kramer, J.H.; Neuhaus, J.; Van Swieten, J.C.; Seelaar, H.; Dopper, E.G.; Onyike, C.U.; et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **2011**, *134*, 2456–2477. <https://doi.org/10.1093/brain/awr179>.
29. McKeith, I.G.; Boeve, B.F.; Dickson, D.W.; Halliday, G.; Taylor, J.P.; Weintraub, D.; Aarsland, D.; Galvin, J.; Attems, J.; Ballard, C.G.; et al. Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* **2017**, *89*, 88–100. <https://doi.org/10.1212/WNL.00000000000004058>.
30. Sachdev, P.; Kalaria, R.; O'Brien, J.; Skoog, I.; Alladi, S.; Black, S.E.; Blacker, D.; Blazer, D.G.; Chen, C.; Chui, H.; et al. Diagnostic criteria for vascular cognitive disorders: a VASCOG statement. *Alzheimer Disease & Associated Disorders* **2014**, *28*, 206–218. <https://doi.org/10.1097/WAD.0000000000000034>.
31. Wilson, S.M.; Galantucci, S.; Tartaglia, M.C.; Gorno-Tempini, M.L. The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language* **2012**, *122*, 190–198. <https://doi.org/10.1016/j.bandl.2012.04.005>.
32. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A Framework for Building Perception Pipelines. *arXiv preprint* **2019**, [arXiv:cs.DC/1906.08172].
33. Russell, J.A. A circumplex model of affect. *Journal of Personality and Social Psychology* **1980**, *39*, 1161–1178. <https://doi.org/10.1037/h0077714>.
34. Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion. *Journal of personality and social psychology* **1971**, *17*, 124–129. <https://doi.org/10.1037/h0030377>.
35. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* **2017**, *10*, 18–31. <https://doi.org/10.1109/TAFFC.2017.2740923>.
36. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
37. Pellegrini, E.; Ballerini, L.; del C. Valdes Hernandez, M.; Chappell, F.M.; González-Castro, V.; Anblagan, D.; Danso, S.; Muñoz-Maniega, S.; Job, D.; Pernet, C.; et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **2018**, *10*, 519–535. <https://doi.org/10.1016/j.dadm.2018.07.004>.
38. Li, J.; Jin, K.; Zhou, D.; Kubota, N.; Ju, Z. Attention mechanism-based CNN for facial expression recognition. *Neurocomputing* **2020**, *411*, 340–350. <https://doi.org/10.1016/j.neucom.2020.06.014>.
39. Huang, Q.; Huang, C.; Wang, X.; Jiang, F. Facial expression recognition with grid-wise attention and visual transformer. *Information Sciences* **2021**, *580*, 35–54. <https://doi.org/10.1016/j.ins.2021.08.043>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.