Multimodal Classification of Alzheimer's Disease by Combining Facial and Eye-Tracking Data

Machine Learning for Health (ML4H) 2024

Shih-Han Chou, Miini Teng, Harshinee Sriram, Chuyuan Li, Giuseppe Carenini, Cristina Conati, Thalia S. Field, Hyeju Jang, Gabriel Murray

Motivation

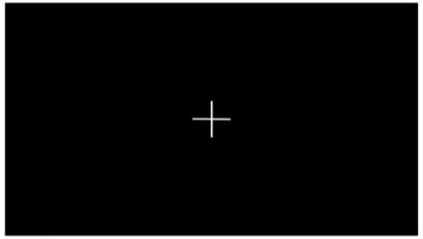
Using existing video data, we explore how facial patterns can enhance the CANARY model for early dementia detection

- **Clinical Correlation**: Apathy, a common symptom in Alzheimer's Disease (AD), leads to reduced goal-directed behavior and flat affect, resulting in decreased facial expressivity (Seidl et al., 2012).
- **Dementia Progression**: Advanced stages of AD impair patients' ability to show appropriate facial emotional reactions, even to emotional stimuli (Asplund et al., 1991).
- **Contrasting Patterns**: While dementia may reduce control over negative expressions, patients may exhibit increased facial expressiveness or use smiles as compensatory behavior in cognitive tasks (Matsushita et al., 2018; Smith, 1995).

Given these nuanced relationships between dementia and facial expressivity, integrating facial analysis could augment cognitive screening, providing a non-invasive, scalable tool for identifying early cognitive impairments.

Dataset

- The dataset comprises data gathered from patients diagnosed with AD or exhibiting initial symptoms potentially progressing to AD.
- Four tasks, including Pupil Calibration, Picture Description, Reading, and Memory. Completing these tasks took an average of 7 minutes.

Instructions:	Visual stimulus:
Pupil calibration task "A cross will appear in the middle of the screen. Please fixate your eyes on the cross. Do not look away from it. This will take about 10 s."	 recall
Picture description task "You will be shown a picture on the screen. Describe everything you see going on in this picture. Try not to look away from the screen while describing the picture."	
Reading task "You will be shown a paragraph on the screen. Please read the paragraph out loud."	In areas where it is very hot and dry, plants and animals have to adapt to these conditions. Many plants survive times of drought in the form of seeds which often lie buried in the ground for several years and do not put out shoots before it rains. When that happens, the plants grow very quickly and form flowers and seeds, which in due time develop into the next generation. Some animals behave in a similar way. There are frogs that bury themselves in the ground and form a capsule which prevents them from drying out. These frogs only come to the surface when it finally rains. They use this time in which water is available to provide for their offspring. A lot of plants in the desert have adapted to the dryness in other ways. Some have extensive roots that take in water from a large area or reach into the ground very far.
Memory description task: "Please recall a positive life event. Some examples are listed here: Your first job, how you met your best friend, a place you have traveled, your favorite teacher, your first pet, or the birth of your first child."	NO VISUAL STIMULUS PROVIDED

Dataset

- The dataset comprises data gathered from patients diagnosed with AD or exhibiting initial symptoms potentially progressing to AD.
- Four tasks, including Pupil Calibration, Picture Description, Reading, and Memory. Completing these tasks took an average of 7 minutes.
- In total, this dataset contains 144 participants where 75 are control and 69 are AD patients.

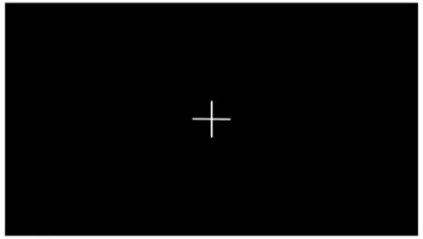
Participant Demographics.
Parti. = Participants, M = Male, F = Female,
MoCA = Montreal Cognitive Assessment Score.

Group	Parti.	Age	Gender	MoCA
Control	75	62 ± 15	22M/53F	27 ± 3
Patients	69	72 ± 9	33M/36F	18 ± 7

Instructions:

Visual stimulus:

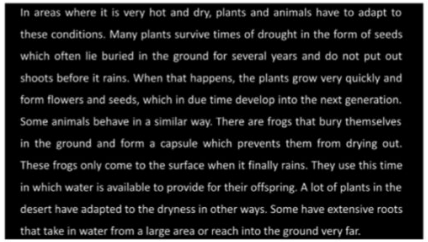
Pupil calibration task
"A cross will appear in the middle of the screen. Please fixate your eyes on the cross. Do not look away from it. This will take about 10 s."



recall

Picture description task
"You will be shown a picture on the screen. Describe everything you see going on in this picture. Try not to look away from the screen while describing the picture."

Reading task
"You will be shown a paragraph on the screen. Please read the paragraph out loud."

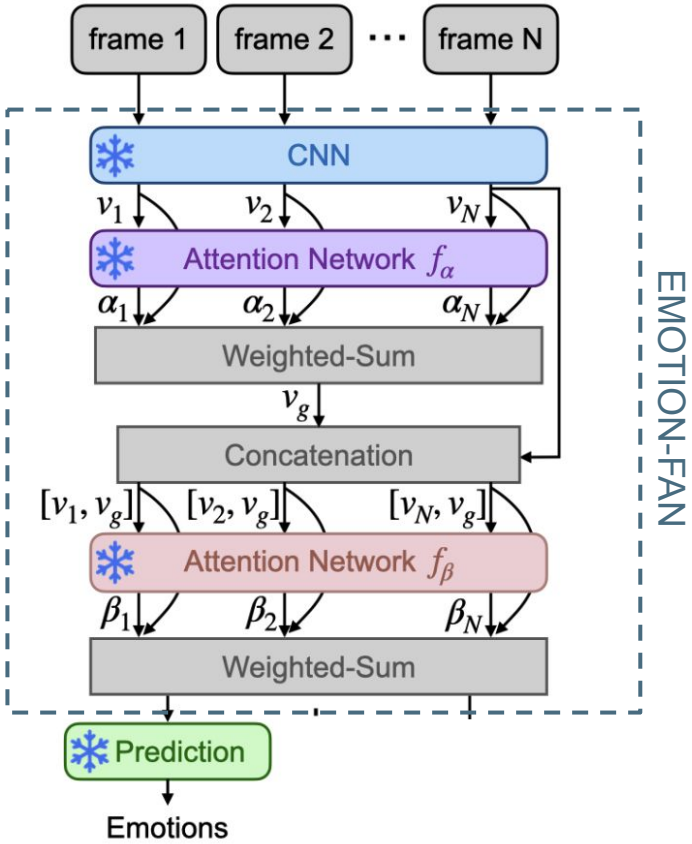


Memory description task:
"Please recall a positive life event. Some examples are listed here: Your first job, how you met your best friend, a place you have traveled, your favorite teacher, your first pet, or the birth of your first child."

NO VISUAL STIMULUS PROVIDED

Approach

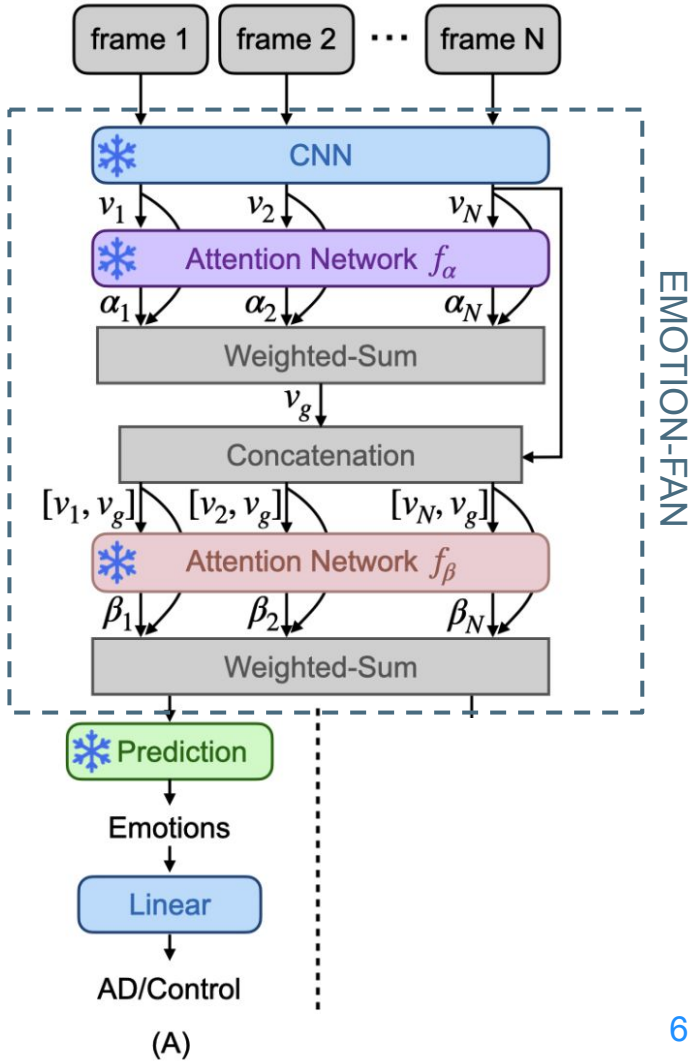
- We build our approach on the top of EMOTION-FAN model. It contains:
  - A CNN network to extract facial representations.
  - Two attention networks to learn frame/video features.
  - A prediction layer for the final prediction.



Approach

- We build our approach on the top of EMOTION-FAN model. It contains:
  - A CNN network to extract facial representations.
  - Two attention networks to learn frame/video features.
  - A prediction layer for the final prediction.
- We use EMOTION-FAN as the backbone and introduce two variations for our AD/health classification task:

  (A) An linear layer is added to explore whether it is feasible to directly detect AD patients based on facial emotions.

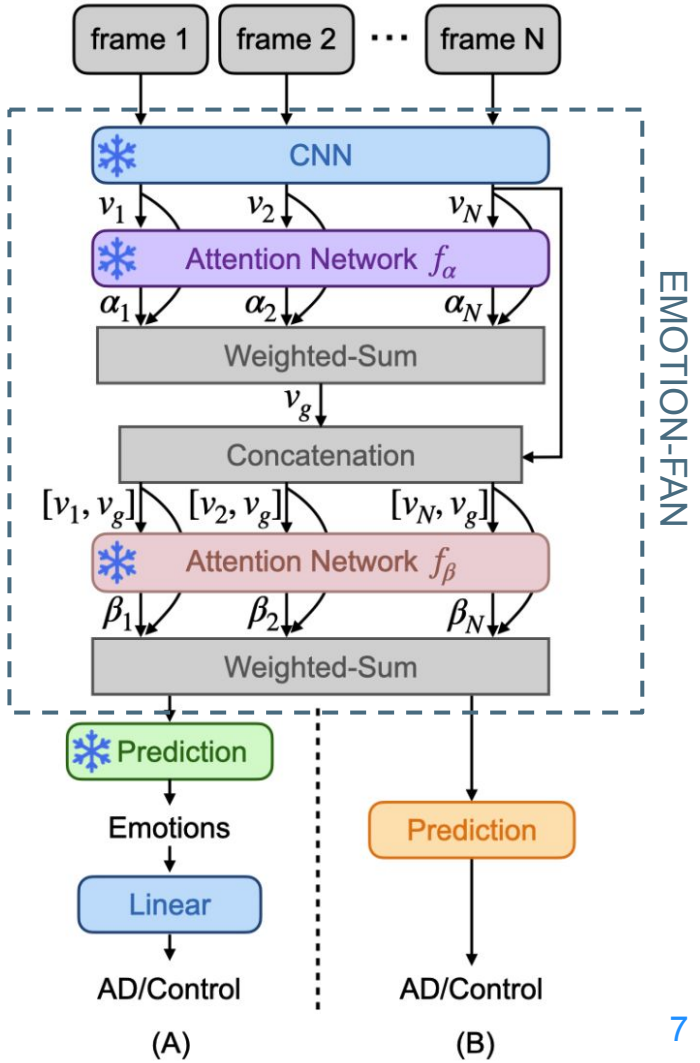


Approach

- We build our approach on the top of EMOTION-FAN model. It contains:
  - A CNN network to extract facial representations.
  - Two attention networks to learn frame/video features.
  - A prediction layer for the final prediction.
- We use EMOTION-FAN as the backbone and introduce two variations for our AD/health classification task:

  (A) An linear layer is added to explore whether it is feasible to directly detect AD patients based on facial emotions.

  (B) We fine-tune the prediction layer to investigate the use of facial embeddings for AD classification.



Results

- AD classification results using facial data on four tasks with two different variations.

Task	Features	Models	AUC	Sensitivity	Specificity
-	*Age (baseline)*	1-layer neural network	0.55 ± 0.05	0.64 ± 0.06	0.63 ± 0.05
Pupil Calibration	Emotions	EMOTION-FAN+linear	0.56 ± 0.03	0.66 ± 0.04	0.65 ± 0.05
	Facial patterns	EMOTION-FAN (ours)	$\mathbf{0.81 \pm 0.02}$	$\mathbf{0.84 \pm 0.03}$	$\mathbf{0.80 \pm 0.04}$
Picture Description	Emotions	EMOTION-FAN+linear	0.57 ± 0.04	0.64 ± 0.06	0.67 ± 0.06
	Facial patterns	EMOTION-FAN (ours)	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.82 \pm 0.03}$	$\mathbf{0.78 \pm 0.03}$
Reading	Emotions	EMOTION-FAN+linear	0.68 ± 0.06	0.70 ± 0.06	0.73 ± 0.05
	Facial patterns	EMOTION-FAN (ours)	$\mathbf{0.83 \pm 0.02}$	$\mathbf{0.83 \pm 0.03}$	$\mathbf{0.81 \pm 0.02}$
Memory	Emotions	EMOTION-FAN+linear	0.61 ± 0.03	0.67 ± 0.05	0.68 ± 0.05
	Facial patterns	EMOTION-FAN (ours)	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.77 \pm 0.03}$	$\mathbf{0.82 \pm 0.03}$

Results

AD Classification Using Facial + Eye-tracking (ET) Data

- Building on the promising results from facial patterns, we investigates the potential of combining facial and ET data for more accurate AD classification.
- We explore a late fusion method which is generally employed for multimodal data. The late fusion method aggregates predictions from ET and video modalities at the decision level.

Task	Modality	Models	AUC	Sensitivity	Specificity
Pupil Calibration	ET	VTNet (Sriram et al., 2023)	0.78 ± 0.01	0.71 ± 0.02	0.75 ± 0.01
	Video	EMOTION-FAN (ours)	0.81 ± 0.02	0.84 ± 0.03	0.80 ± 0.04
	ET+Video	VTNet + EMOTION-FAN (ours)	$\mathbf{0.84 \pm 0.02}$	$\mathbf{0.85 \pm 0.04}$	$\mathbf{0.81 \pm 0.03}$
Picture Description	ET	VTNet (Sriram et al., 2023)	0.76 ± 0.01	0.70 ± 0.02	0.73 ± 0.02
	Video	EMOTION-FAN (ours)	0.79 ± 0.02	0.82 ± 0.03	0.78 ± 0.03
	ET+Video	VTNet + EMOTION-FAN (ours)	$\mathbf{0.83 \pm 0.02}$	$\mathbf{0.82 \pm 0.03}$	$\mathbf{0.82 \pm 0.03}$
Reading	ET	VTNet (Sriram et al., 2023)	0.78 ± 0.01	0.70 ± 0.01	0.80 ± 0.02
	Video	EMOTION-FAN (ours)	0.83 ± 0.02	0.83 ± 0.03	0.81 ± 0.02
	ET+Video	VTNet + EMOTION-FAN (ours)	$\mathbf{0.88 \pm 0.01}$	$\mathbf{0.86 \pm 0.03}$	$\mathbf{0.86 \pm 0.03}$
Combined†	ET+Video	VTNet + EMOTION-FAN (ours)	0.88 ± 0.01	0.85 ± 0.03	0.86 ± 0.03

Take-away messages

- Fine-tuned EMOTION-FAN model greatly outperforms the model using eye-tracking data, validating facial patterns provide valuable information.
- Late fusion strategy combining eye-tracking and video data shows the benefits of multimodal data integration.
- Detailed analysis, such as,
  - MANOVA results
  - Confusion Matrices
  - LIME visualization

  are provided in the main paper.

Please see our paper or visit our poster session (location: 49) for more information.

THE UNIVERSITY OF BRITISH COLUMBIA

Thank you and welcome to our poster (location: 49) at 14:15-15:15
if you have any question or would like to learn more!