Leveraging Eye Movement Features for Enhanced Emotion Recognition Accuracy

Tajudin Abdela Ademe[1] Ritika Jain[2] Nirav P Bhatt[3]

Indian Institute of Technology Madras (IIT Madras)

INTRODUCTION

- Emotion recognition is vital in affective computing and HCI, and eye movement has emerged as a promising modality.
- Prior work shows that blink patterns and pupil changes correlate with emotional states.
- Using eye data alone can offer a non-invasive, cost-effective alternative to EEG.
- This work proposes a deep model (DGCCA-AM) to classify emotions from eye movement alone.

PROBLEM STATEMENT

- Most existing emotion classifiers rely on EEG or multimodal inputs, while purely eye-based methods remain under-explored.
- Standalone eye-tracking systems often lack sophisticated feature fusion.
- There is a need for models that explicitly group and fuse eye movement features (e.g. pupil, saccades, events) to fully exploit their emotional cues.

OBJECTIVES

- Develop an interpretable deep model (DGCCA-AM) using only eye-tracking data for four-class emotion classification. We group eye features into semantically meaningful sets and apply DGCCA with an attention mechanism to align and fuse them for robust emotion prediction.

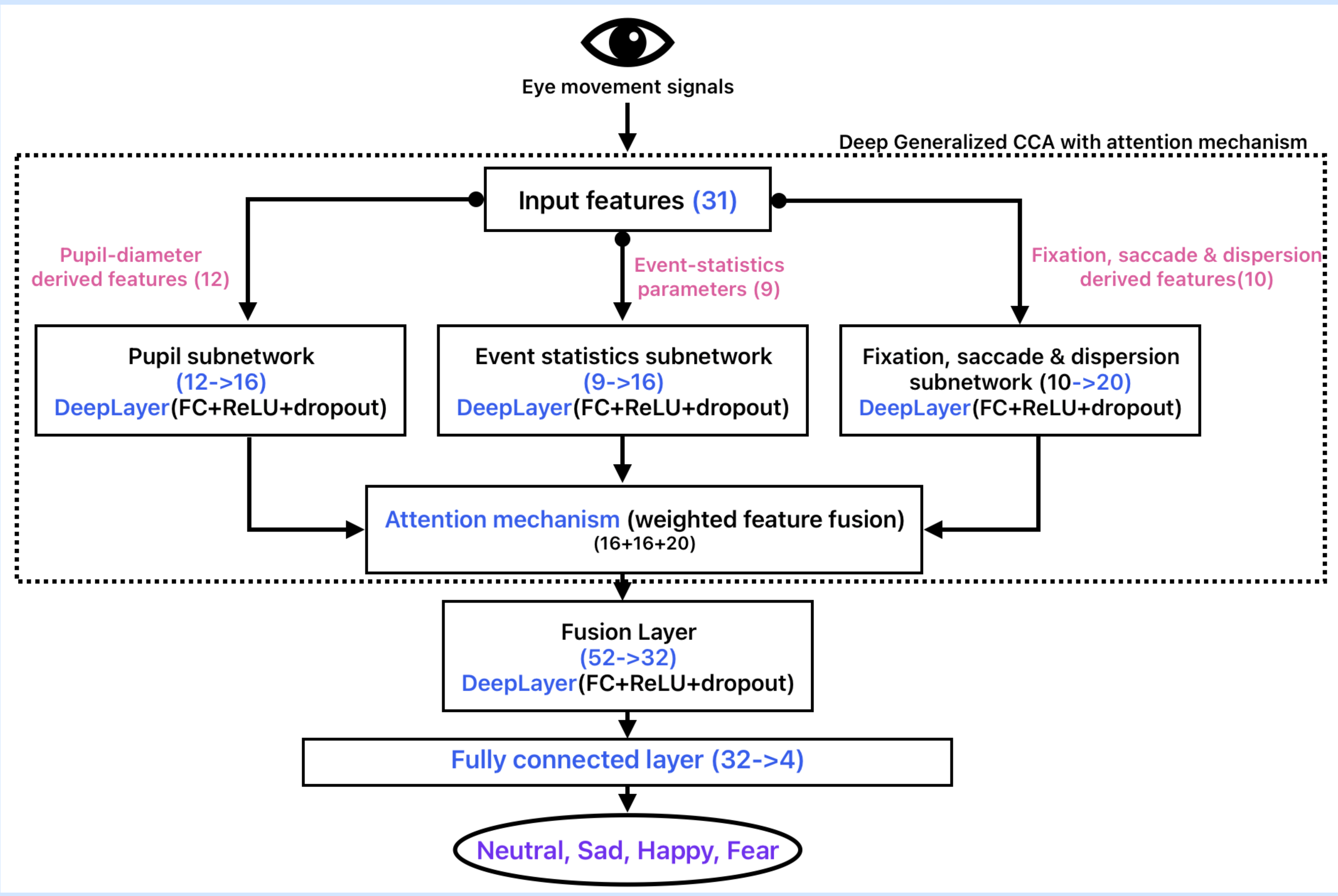
DATASETS

Class distribution in training, validation, and test sets for Session 3 of the SEED-IV dataset.

Emotion	Training	Validation	Test
Neutral	2473	275	687
Sad	2462	274	677
Fear	1760	196	479
Happy	2182	242	602

PROPOSED METHOD

To leverage semantic structure in eye movement data, we divided features into three semantic groups (31 features total):

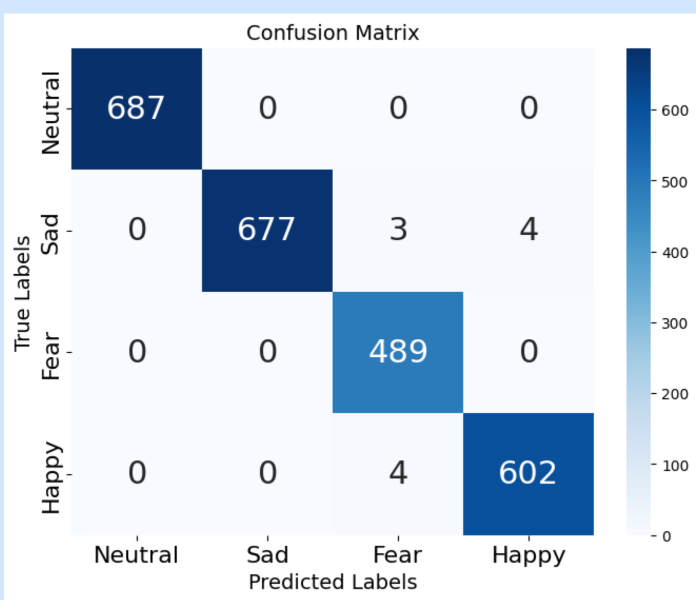


- **Pupil-Based Features (12)**:Capture emotional arousal via pupil diameter stats (mean, std. for both eyes) and Differential Entropy (DE) across 4 frequency bands: 0–0.2, 0.2–0.4, 0.4–0.6, and 0.6–1.0 Hz.
- **Fixation, Saccade Dispersion Features (10)**: Model gaze control and attention using saccade amplitude/duration (mean, std.), fixation duration, and gaze dispersion along X and Y axes.
- **Event-Based Features (9)**: Quantify blink rate, fixation/saccade counts and durations, average saccade amplitude, and latency to capture temporal gaze event patterns.

Deep Generalized Canonical Correlation Analysis with Attention Mechanism (DGCCA-AM) model structure.

- **Group-Specific Subnetworks: Each feature group (Pupil, Fixation/Saccade, Events)** is passed through a separate subnetwork with a fully connected (FC) layer and ReLU activation to learn group-specific embeddings.
  Each group X_i is passed through a dedicated subnetwork:
  $$X_i \to \mathrm{FC}_i \to \mathrm{ReLU} \to h_i$$
  producing latent vectors h_1, h_2, h_3.
- **DGCCA Alignment Layer:** Learns shared latent representations by maximizing correlation between the feature groups using Deep Generalized Canonical Correlation Analysis (DGCCA).
  $$\max_{W_1, W_2, \ldots, W_n} \sum_{i<j} \mathrm{corr}(W_i X_i, W_j X_j)$$
- **Attention Fusion Mechanism:** An attention layer computes weights for each aligned feature group, dynamically emphasizing the most informative signals per sample.
  $$\mathbf{z} = \sum_{i=1}^{3} \alpha_i h_i, \quad \text{where} \quad \alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)}$$
- Incorporating DGCCA (Deep Generalized CCA) encourages each subnetwork (modality) to learn representations that are maximally correlated across the different modalities.
- Weighted features from each modality are concatenated and passed through a fully-connected fusion layer that reduces the feature dimension from 52 to 32. The resulting fused vector is then fed into a Softmax classifier (32→4) to predict one of four emotion classes (Happy, Sad, Neutral, Fear).
- The model is trained end-to-end with the Adam optimizer and a categorical cross-entropy loss function.
- Performance is evaluated under both subject-dependent and subject-independent settings, using accuracy scores and confusion matrices to measure classification effectiveness.

RESULTS



- Session 3 (intra-subject) confusion matrix shows minimal misclassifications.
- Intra-subject accuracy peaked at 99.92
- Feature ablation: baseline (all features) accuracy 99.92%; without pupil features 93.71% (–6.2%); without event-statistics 97.77% (–2.1%); without fixation/saccade 97.93% (–2.0%).

Subject ID	Session 1	Session 2	Session 3	Avg. across sessions
1	41.72	68.63	69.71	60.02
2	53.82	57.45	94.28	68.52
3	59.81	65.02	65.21	63.35
4	35.02	47.24	54.99	45.75
5	72.50	74.52	60.46	69.16
6	60.05	52.16	84.43	65.55
7	72.74	61.18	73.72	69.21
8	64.39	59.01	74.57	65.99
9	62.87	75.96	72.99	70.61
10	43.95	71.15	77.98	64.36
11	76.85	56.13	42.94	58.64
12	57.46	61.54	53.16	57.39
13	45.95	72.48	76.40	64.94
14	79.55	58.65	77.74	71.98
15	50.65	65.87	38.32	51.61
Average	**58.49**	**63.13**	**67.79**	**63.14**

- Inter-subject (LOSO) overall average accuracy 63.14% across all subjects and sessions.
- Session 3 (inter-subject) average accuracy 67.79% across subjects (highest of the three sessions).

CONCLUSION

- We demonstrate that eye movement features alone can achieve near-perfect emotion classification in a subject-dependent setting (99.92% accuracy)
- The DGCCA-AM model effectively fuses semantically grouped features to capture emotion-relevant signals.
- A key finding is the dominant role of pupil-based features and event statistics in recognition performance.
- Eye data alone can rival multimodal methods when models are tailored to them. However, cross-subject generalization remains limited (63% accuracy), underscoring the need for subject-specific or adaptive techniques.

LIMITATION AND FUTURE WORK

- Limited cross-subject generalization due to individual variability in eye movement patterns.
- Domain adaptation methods needed for robust, subject-invariant emotion recognition.
- Future work: evaluate in diverse, real-world settings for practical deployment.

REFERENCES

Lan, Y.-T., W. Liu, and B.-L. Lu (July 2020). "Multimodal emotion recognition using deep generalized canonical correlation analysis with an attention mechanism". In: *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, UK, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207625. URL: https://doi.org/10.1109/IJCNN48605.

tajudinabdela4@gmail.com