Detect Mild Cognitive Impairment in older adults using facial videos

Jian Sun[1], Hiroko H. Dodge[2], Mohammad H. Mahoor[3]

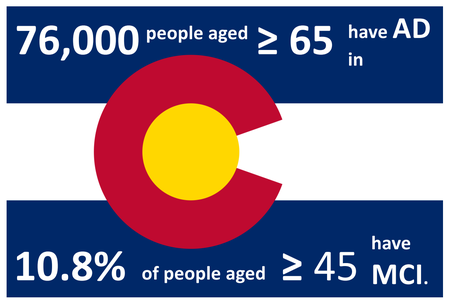
[1]Department of Computer Science, University of Denver. [2]Department of Neurology at Harvard Medical School, Harvard University.
[3]Department of Electrical and Computer Engineering, University of Denver.

What we do?

- **We develop deep learning model to detect MCI (Mild Cognitive Impairment) from NC (Normal Cognition) at the early stage on causal chat video.**

Why?

- **MCI ☞ Alzheimer's disease (AD), a public health crisis in CO.**

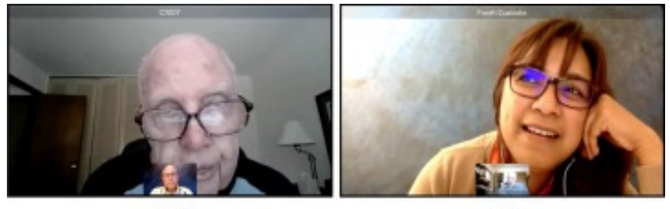


Detecting MCI at the early stage matters!

- **Common datasets are hard to access and co$tly (MRI).**
- **Current models are weak at getting sufficient or temporal features and computational cost (SVM, CNN, 3D-CNN).**

Dataset

- **I- CONECT(Internet-Based Conversational Engagement Clinical Trial) comprises video records of conversations.**



(a) Interviewee is talking; (b) interviewer is speaking.

- **186 people (aged ≥ 75, 100 MCI, 86 NC) from the USA .**
- **161 themes | 30 min/session | 4/week | > 6 months.**

Problem

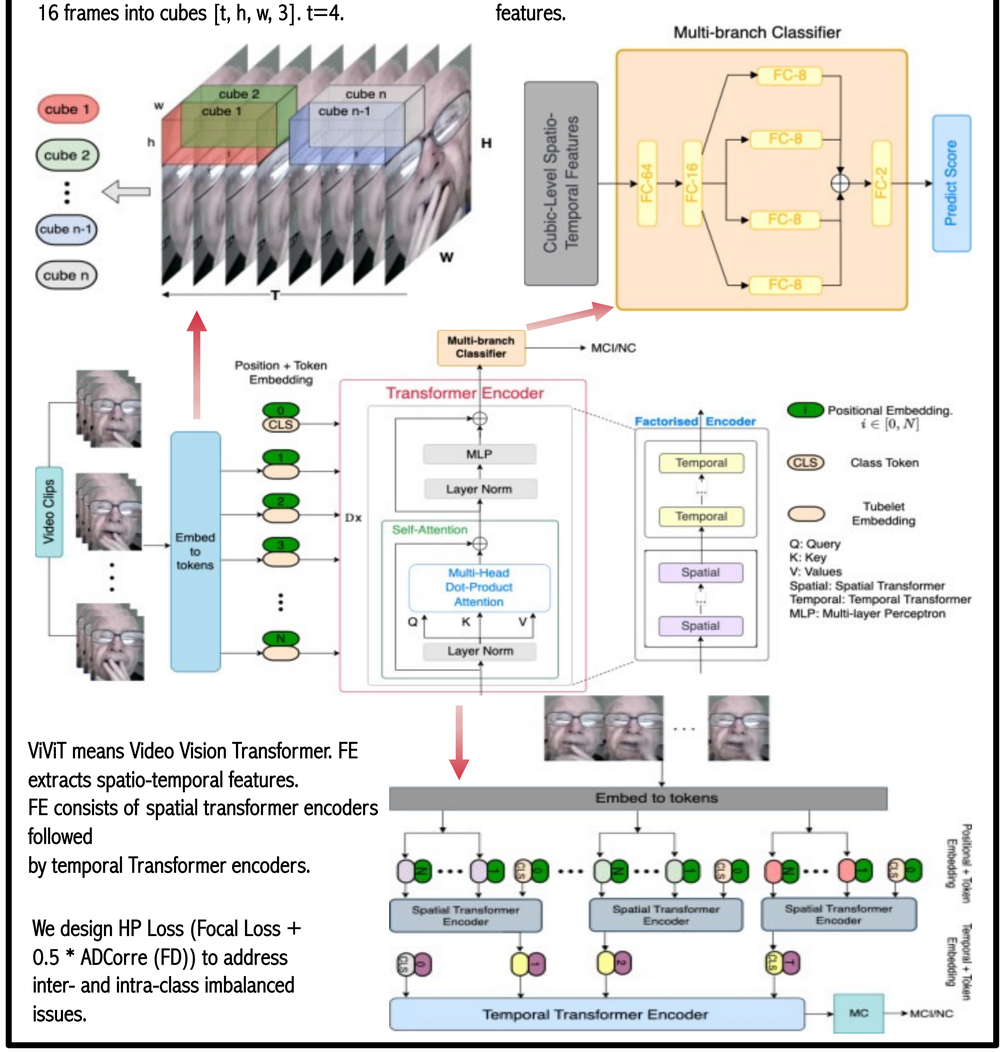
- **Inter-class Imbalance: more MCIs than NCs.**
- **Intra-class Imbalance: various video length means different feature volumes.**

MC-ViViT

- **The new proposed MC-ViViT contains ViViT backbone with Factorized Encoder (FE) and Multi-branch Classifier (MC).**

MC-ViViT applies tubelet embedding to cut consecutive 16 frames into cubes [t, h, w, 3]. t=4.

MC helps view object from different perspective and enrich features.

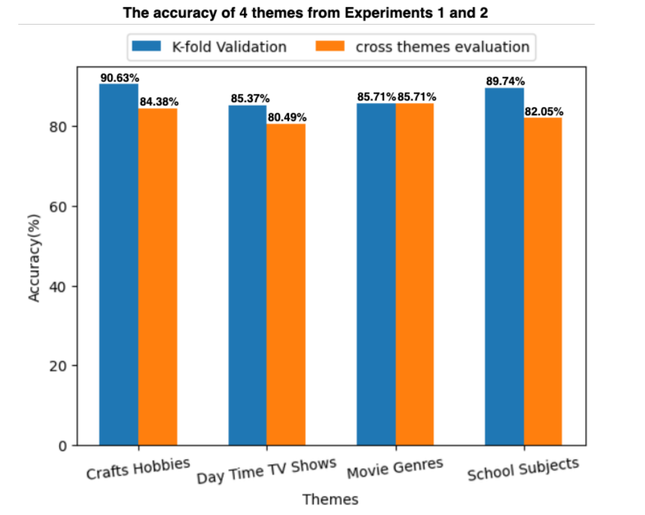


ViViT means Video Vision Transformer. FE extracts spatio-temporal features.
FE consists of spatial transformer encoders followed
by temporal Transformer encoders.

We design HP Loss (Focal Loss + 0.5 * ADCorre (FD)) to address inter- and intra-class imbalanced issues.

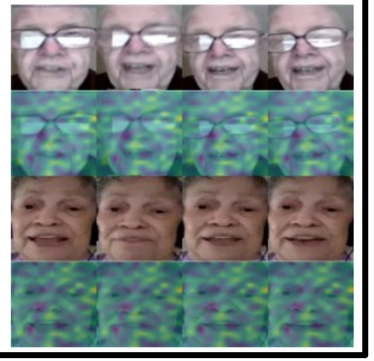
Results

- **Experiment 1: Train and test on the certain theme.**
- **Here, MC-ViViT gets 90.63% accuracy on the Crafts Hobbies.**
- **Experiment 2: Train on 3 themes and test on one.**
- **MC-ViViT gets good results on both experiments, which shows the generalization of the MC-ViViT on I-CONECT dataset.**
- **The HP Loss addresses the inter- and intra-class imbalance issues well.**



Conclusions

- **MC-ViViT detects MCI via facial videos of the semi-structured interviews well.**
- **MC-ViViT pays attention to the different areas of the face (forehead, eyelids, nose, cheek, and jaw).**



Acknowledgement

Colorado Office of Economic Development; International Trade and National Institutes of Health (NIH) (R01AG051628 and R01AG056102).