MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Speech Emotion Recognition Applied to Real-World Medical Consultation

Ching-Tzu Huang, Chih-Wei Huang, Hsuan-Chia Yang*, and Yu-Chuan Jack Li*

Graduate Student

Graduate Institute of Biomedical Informatics, Taipei Medical University, Taiwan

@Janicexice

minijennis@gmail.com

AIDH IMIA

Introduction

- A **satisfying cognitive interaction** led to increased physician trust and expertise perception.
- Surgeons who have been sued are judged more **dominant and less concerned** in their **voice**.(non-verbal)
- Masks cover **emotional transmission**.

Image by pressfoto on Freepik

Introduction

- What is **"Emotions"?**

Valence

Negative ⟷ Positive

Arousal

Calm ⟷ Excited

Dominance

Low ⟷ High

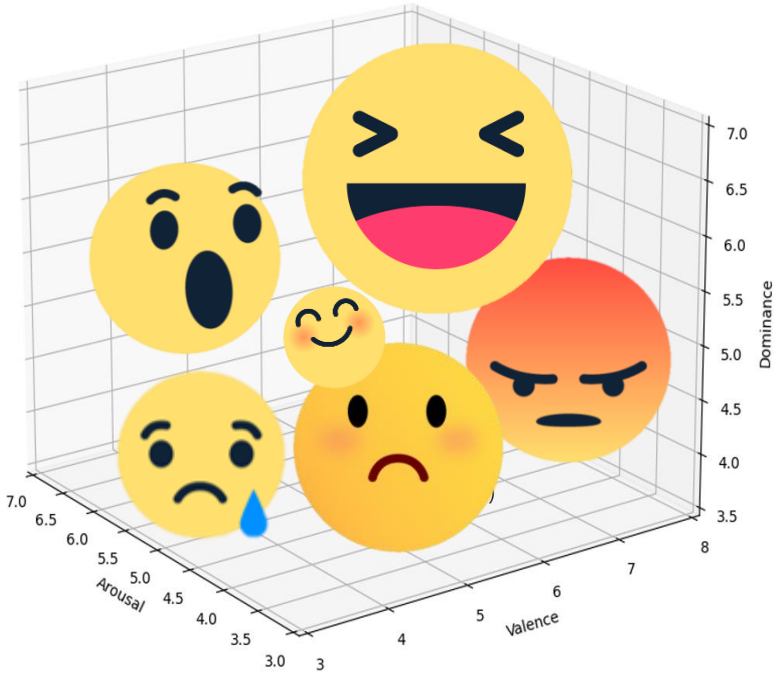
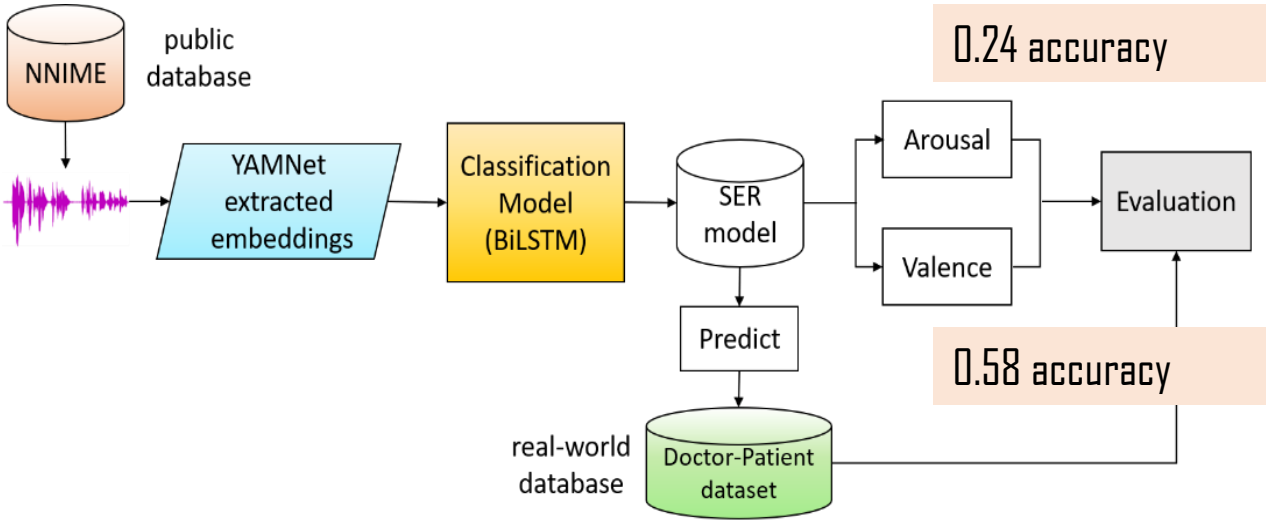


Image by Freepik

Study Objectives

- a **deep learning** model specialized in **doctor-patient communication**
- Education in **medical simulation**

Acting dataset→Real clinical recordings



Doctor-Patient interaction Corpus

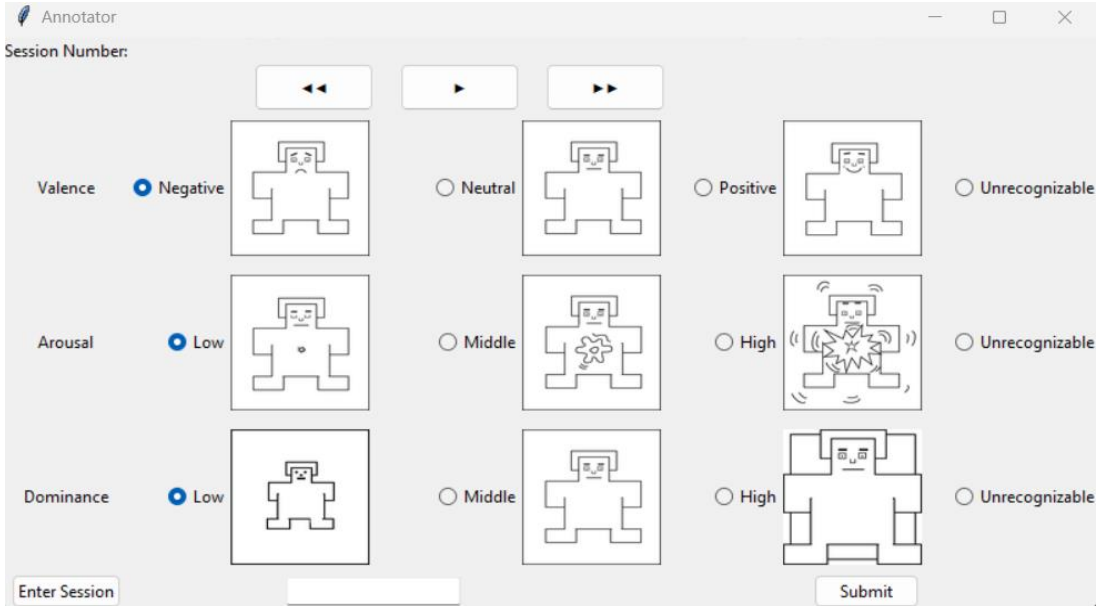
- Dermatology clinic room
- **4 doctors**(2 females, 2 males)
- **304 conversations** (different patients)
- 11,264 segments
- Average 4.67 s

(the Doctor's speech)

(the Patient's speech)

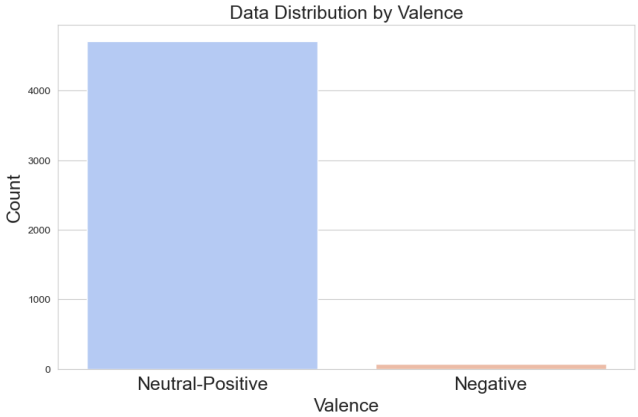
Doctor-Patient interaction Corpus

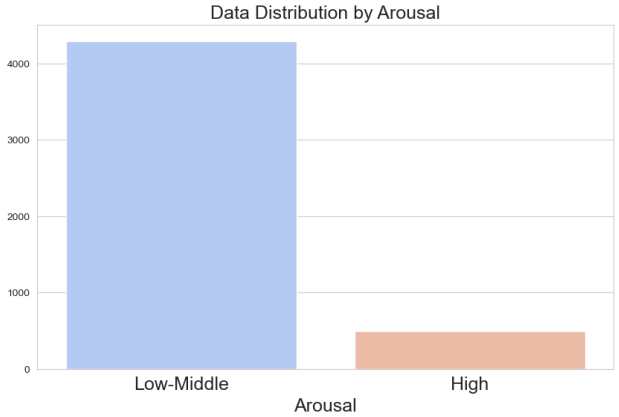
- 3 categories
- 304 sessions
- 3 annotators

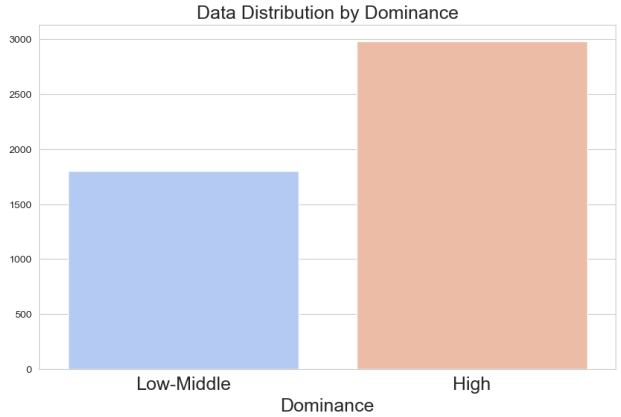


Doctor-Patient interaction Corpus

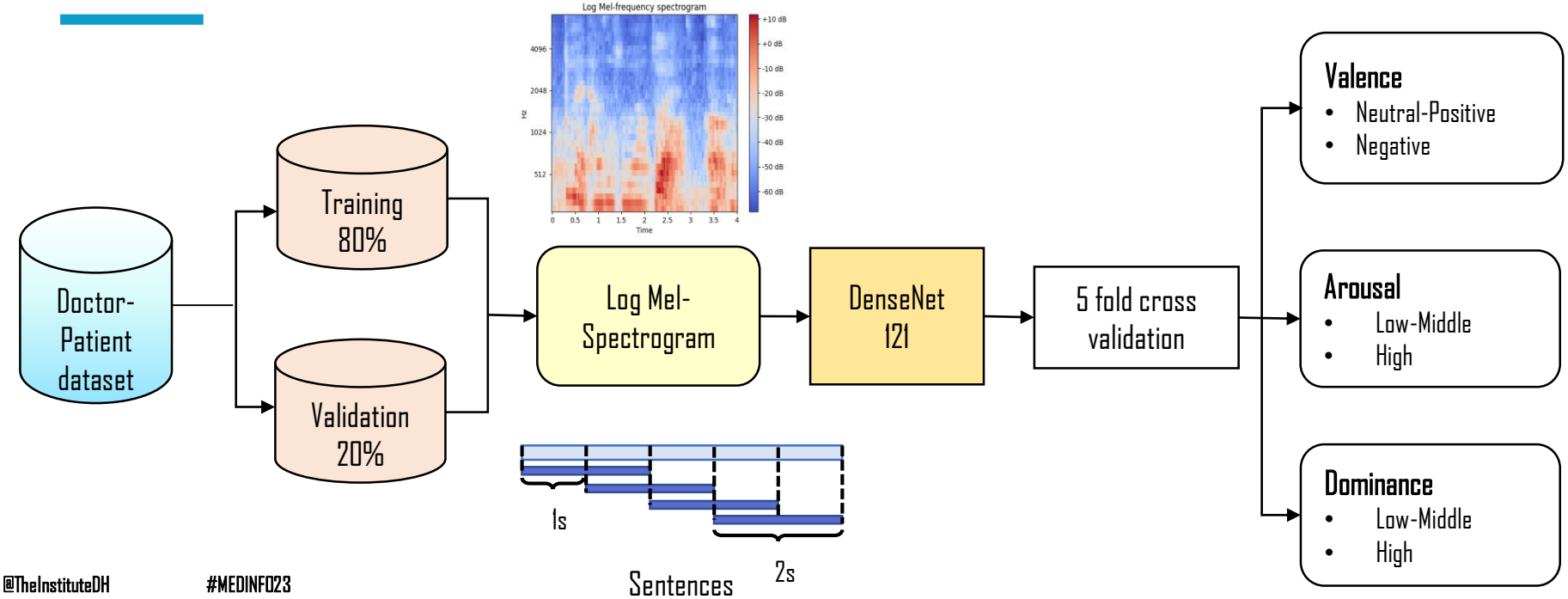
- Data Imbalance problem
- 3-class⟶2-class







Doctor-Patient emotion prediction



Speech emotion prediction - DenseNet121

Model	Duration	Validation AUROC
Valence	4s	0.76
	5s	0.73
	6s	0.82
Arousal	4s	0.81
	5s	0.80
	6s	0.81

Model	Duration	Validation AUROC
Dominance	4s	0.75
	5s	0.77
	6s	0.77

Speech emotion prediction – Testing 5281 segments

Model	Accuracy	Testing AUROC	Sensitivity	Specificity	Precision	Recall
Valence	0.54	0.51	0.55	0.54	0.01	0.55
Arousal	0.68	0.84	0.85	0.67	0.18	0.85
Dominance	0.69	0.76	0.69	0.69	0.77	0.69

*At the optimal threshold

Discussion

- Human annotation cannot be completely separated from semantics
- Predicted length of time
- Obvious emotions are scarce
- Speaker diversity and multilanguage

Reference

- Khan, F. H., Hanif, R., Tabassum, R., Qidwai, W., & Nanji, K. (2014). Patient attitudes towards physician nonverbal behaviors during consultancy: result from a developing country. International Scholarly Research Notices, 2014..
- Kim, S. S., Kaplowitz, S., & Johnston, M. V. (2004). The effects of physician empathy on patient satisfaction and compliance. Evaluation & the health professions, 27(3), 237-251.
- Mehrabian, A. (2017). Nonverbal communication: Routledge.
- Ranjan, P., Kumari, A., & Chakrawarty, A. (2015). How can doctors improve their communication skills?. Journal of clinical and diagnostic research: JCDR, 9(3), JE01.
- Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.
- Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. Communications of the ACM, 61(5), 90-99..
- Ye, J., Wen, X., Wei, Y., Xu, Y., Liu, K., & Shan, H. (2022). Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech Emotion Recognition. arXiv preprint arXiv:2211.08233.
- Zhang, H., Gou, R., Shang, J., Shen, F., Wu, Y., & Dai, G. (2021). Pre-trained deep convolution neural network model with attention for speech emotion recognition. Frontiers in Physiology, 12, 643202.
- Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical signal processing and control, 47, 312-323.

Reference

- 蕭隆城. (2007). 醫師情緒智能, 醫病關係與醫療糾紛之相關性探討─ 以南部某區域醫院為例.
- Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication, 116, 56-76.
- Kehrein, R. (2002). The prosody of authentic emotions. In Speech Prosody 2002, International Conference.
- Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: a clue to malpractice history. *Surgery, 132*(1), 5-9.
- Bestelmeyer, P. E., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social cognitive and affective neuroscience, 12*(8), 1351-1358.
- Bhangale, K., & Mohanaprasad, K. (2022). Speech emotion recognition using mel frequency log spectrogram and deep convolutional neural network. In *Futuristic Communication and Network Technologies* (pp. 241-250): Springer.
- Chou, H.-C., Lin, W.-C., Chang, L.-C., Li, C.-C., Ma, H.-P., & Lee, C.-C. (2017). *Nnime: The nthu-ntua chinese interactive multimodal emotion corpus.* Paper presented at the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII).
- Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech communication, 40*(1-2), 5-32.

Acknowledgements

& ICHIT Lab

Prof.Jack Li

Prof.Edward Yang

Prof.Thomas Lin

Annisa Ristya Rahmanti

Dr.Grace Huang

Prof. Hung-Wen Chiu