Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Early Career Spotlight Track

Human-Centred Trustworthy AI for Digital Health: Sensing, Understanding, and Empowerment

Zhao Ren
Cognitive Systems Lab, University of Bremen, Germany
zren@uni-bremen.de

Abstract

The rapid advancement of Artificial Intelligence (AI) has accelerated the development of personalised healthcare. However, the clinical adoption of deep learning remains constrained by a persistent "trust gap" surrounding model transparency, security, and data privacy. This talk presents a research vision for building trustworthy human-centred AI in computer audition and biosignal processing by integrating signal processing, machine learning, and healthcare. This vision is structured around three interconnected pillars: Sensing, Understanding, and Empowerment. The Sensing pillar focuses on enabling machines to perceive clinically relevant information from biosignals, transforming body sounds and physiological signals into non-invasive, accessible, and cost-effective windows into human health and well-being. The Understanding pillar addresses the black-box nature of modern AI systems, with the goal of ensuring transparency, efficiency, robustness, and security in clinical deployment. This is achieved through the development of explainable AI methods, knowledge distillation techniques, and defences against adversarial attacks, fostering AI systems that clinicians can trust and interpret. The Empowerment pillar seeks to restore natural communication for individuals with speech impairments, such as laryngectomy patients. By developing Silent Speech Interfaces (SSIs) that translate facial muscle activity (EMG) directly into audible speech, this research advances speech intelligibility, naturalness, real-time causal architectures, and multi-speaker speech synthesis. Collectively, these research directions aim to establish a foundation of trustworthy, human-centred AI that is not only accurate and secure, but also empathetic, accessible, and impactful, ultimately enhancing healthcare, communication, and quality of life.

1 Introduction

Digital health is undergoing a profound transformation driven by recent advances in Artificial Intelligence (AI) [Yeung *et al.*, 2023]. Rather than relying on one-size-fits-all approaches, AI enables healthcare solutions that adapt to the characteristics and needs of individual patients. This progress is made possible by the ability of AI to learn from heterogeneous data sources, ranging from structured clinical information (e. g., patient demographics and medical records) [Juhn and Liu, 2020] to unstructured modalities such as medical images, clinical narratives, speech, and physiological signals [Xu *et al.*, 2026; Jiang *et al.*, 2024]. For instance, among the many data modalities used in digital health, speech, audio, and physiological biosignals offer unique advantages. They can be acquired non-invasively, frequently, and often using low-cost wearable or consumer devices, making them particularly attractive for continuous monitoring of physical, cognitive, and mental health. Recent studies have demonstrated their potential for detecting neurological disorders [Cummins *et al.*, 2020], cardiovascular diseases [Ren *et al.*, 2024a], respiratory conditions [Chambres *et al.*, 2018], emotional well-being [Schuller and Batliner, 2013], and other health-related states. At the same time, the widespread adoption of wearable technologies, including smart watches and smart glasses, has created unprecedented opportunities for continuous, real-world health monitoring from these biosignals. By integrating these diverse data streams, AI is paving the way for more proactive, personalised, and data-driven healthcare, with applications spanning disease screening, diagnosis, patient monitoring, treatment planning, and clinical decision support.

While AI has demonstrated remarkable potential in digital health, its translation into routine clinical practice remains constrained by concerns over reliability, transparency, and user trust. Clinical decisions often involve high-stakes scenarios where inaccurate, biased, or poorly understood predictions can have serious consequences. Unlike many consumer AI applications, errors in healthcare may directly affect patient safety and clinical decision-making. Consequently, healthcare AI should satisfy stricter requirements for reliability, robustness, explainability, privacy protection, and fairness before it can be trusted in real-world practice. Moreover, the growing complexity of deep learning models, together with the diverse and noisy nature of healthcare data, makes it increasingly difficult to understand how these systems arrive at their conclusions. Recent advances [Gerczuk *et al.*, 2023; Zhang *et al.*, 2026] in foundation models and multimodal

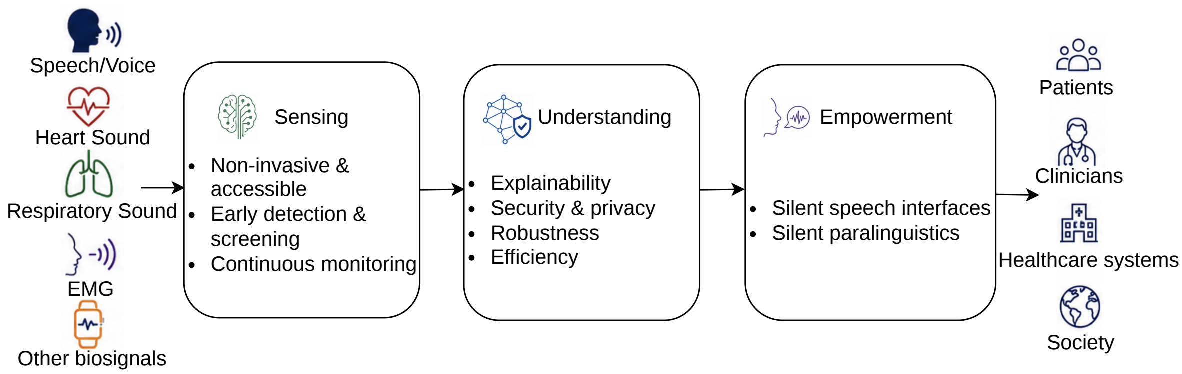

Figure 1: Overview of the proposed research framework. The three pillars, Sensing, Understanding, and Empowerment, provide a roadmap towards trustworthy AI that bridges sensing technologies and future human-centred digital health applications.

large language models are creating unprecedented opportunities for digital health by enabling knowledge transfer across diverse data modalities and clinical tasks. However, their increasing complexity further amplifies the need for trustworthy AI principles. Addressing these challenges requires AI that is not only accurate, but also interpretable, robust, fair, privacy-preserving, and secure. Building such trustworthy AI systems is essential for ensuring safe clinical adoption and fostering confidence among healthcare professionals and patients.

The overarching objective is to advance trustworthy AI for human-centred digital health by developing intelligent systems that transform biosignals into reliable, explainable, and actionable health insights through non-invasive, accessible, and affordable sensing technologies. Rather than replacing clinicians, our goal is to develop AI systems that augment human expertise by providing reliable evidence, transparent reasoning, and actionable recommendations while keeping clinicians and patients at the centre of the decision-making process. Our long-term vision is to bridge sensing technologies and clinical decision support, enabling AI systems that are transparent, robust, secure, and human-centred, thereby supporting informed clinical decisions and improving patient care. This paper presents a research vision for trustworthy AI in digital health through the lens of the authors' research. Rather than surveying the entire field, we organise our contributions into a unified framework consisting of three complementary pillars: *Sensing*, *Understanding*, and *Empowering* (see Figure 1). Together, these three themes provide a roadmap toward human-centred AI systems that are clinically reliable, widely accessible, and capable of improving healthcare outcomes across diverse real-world settings.

2 Sensing

The first pillar of our research vision focuses on enabling AI systems to perceive clinically relevant information from biosignals. Advances in sensing technologies have made it possible to acquire diverse physiological and behavioural signals non-invasively and continuously using wearable and consumer devices. These biosignals provide rich biomarkers of human health, supporting applications ranging from disease screening and diagnosis to continuous monitoring. In the following, we highlight our research on speech for mental healthcare and body sounds for cardiovascular and respiratory health.

2.1 Speech-based Mental Healthcare

Speech is a rich source of information that conveys not only *what* is spoken through linguistic content but also *how* it is spoken through paralinguistic cues. These cues capture long-term characteristics, such as personality, age, and gender, as well as short-term affective states, including emotion [Schuller and Batliner, 2013]. Recent advances in AI, ranging from conventional machine learning approaches to self-supervised speech foundation models and large speech-language models, have substantially advanced automatic speech emotion recognition [Schuller and Batliner, 2013; Ren *et al.*, 2023a; Peng *et al.*, 2024].

Beyond emotion recognition, speech has emerged as a promising digital biomarker for mental and neurological health. In our research, speech has been investigated for recognising mental disorders, including depression [Ringeval *et al.*, 2019] and bipolar disorder [Ren *et al.*, 2019], as well as neurological disorders such as Alzheimer's disease [Cummins *et al.*, 2020]. These studies demonstrate the potential of speech-based AI to support early screening and objective assessment of cognitive and mental health conditions through non-invasive and accessible sensing.

2.2 Body Sounds for Cardiovascular and Respiratory Health

Compared with physiological biosignals such as electrocardiograms (ECGs), body sounds offer a non-invasive, low-cost, and easily accessible means of assessing human health. Our research explores the "acoustic dissection" of body sounds, demonstrating that heart and respiratory sounds contain rich biomarkers for cardiovascular and respiratory conditions [Ren *et al.*, 2024a].

Our work includes the development and benchmarking of the Heart Sounds Shenzhen (HSS) Corpus, which has provided a benchmark for AI-based cardiovascular health monitoring [Dong *et al.*, 2019; Schuller *et al.*, 2018]. Beyond conventional abnormality detection, we investigated automated assessment of heart disease severity using the HSS corpus. For respiratory sound analysis, our research has focused on improving the explainability of AI models to enhance their transparency and facilitate clinical trust. These explainable AI approaches are discussed in detail in Section 3.1.

3 Understanding

The second pillar of our research vision addresses the "black-box" nature of AI by developing models that are transparent, robust, secure, privacy-preserving, and efficient for trustworthy clinical deployment.

3.1 Explainability

The complexity of deep learning models often limits their interpretability, particularly in healthcare applications where transparent decision-making is essential. Our research therefore focuses on explainable AI to improve the understanding of model predictions.

Our early work investigated local explanations using attention mechanisms that highlight clinically relevant time–frequency regions for heart sound analysis, providing intuitive visualisations of model decisions [Ren *et al.*, 2022]. We subsequently extended this idea to global explanations through prototype and criticism learning for respiratory sound analysis [Ren *et al.*, 2023b; Chang *et al.*, 2022]. Instead of explaining individual predictions, these methods identify representative and atypical examples, enabling clinicians to better understand model behaviour and compare new cases with clinically meaningful reference samples.

More recently, we have explored concept-based explanations by linking intermediate representations of speech foundation models to physically meaningful acoustic cues [Rammohan *et al.*, 2025]. We further validated explainable AI methods by analysing the acoustic characteristics of regions highlighted by explanation models [Nasr *et al.*, 2026]. Together, these studies advance explainable AI from prediction-level interpretation towards a more comprehensive understanding of learnt representations.

3.2 Security and Privacy

Ensuring the security and privacy of AI systems is essential for trustworthy healthcare applications. Deep learning models are vulnerable to adversarial attacks, where imperceptible perturbations can cause incorrect predictions, potentially leading to inappropriate clinical decisions. Our research has therefore investigated both the vulnerability of AI models and effective defence strategies.

On the offensive side, we developed methods to improve the transferability and sparsity of adversarial attacks, including lifelong learning-based attack generation [Ren *et al.*, 2020b] and the STAA-Net framework [Chang *et al.*, 2024]. On the defensive side, we proposed adversarial training that improves model robustness by jointly optimising model predictions and intermediate representations using adversarial samples [Ren *et al.*, 2020a]. More recently, we extended these ideas to federated learning and introduced a two-stage defence combining adversarial training with inference-time randomisation [Chang *et al.*, 2026].

Beyond model robustness, our research emphasises privacy-preserving AI for healthcare. We investigated federated learning to enable collaborative model training without sharing sensitive patient data [Chang *et al.*, 2026], and contributed a comprehensive survey on machine unlearning and privacy-preserving explanations, highlighting approaches for removing sensitive information from trained models while reducing the risk of data leakage [Nguyen *et al.*, 2025].

3.3 Model Efficiency

Efficient AI models are essential for deploying digital health applications on resource-constrained devices, such as smartphones and wearable sensors. Our research has therefore focused on improving model efficiency while maintaining high predictive performance.

We proposed a self-distillation framework that transfers knowledge from large speech foundation models to compact student models, enabling efficient speech emotion recognition with minimal performance degradation [Ren *et al.*, 2023a]. To further reduce storage requirements, we investigated dataset distillation, generating compact synthetic datasets that preserve representative characteristics of the original data while reducing memory consumption and mitigating privacy risks [Chang *et al.*, 2025].

In addition, we developed lightweight neural architectures based on parameterised hypercomplex neural networks to reduce model complexity for physiological signal analysis [Basso *et al.*, 2023]. These studies facilitate real-time AI on edge devices, bringing trustworthy digital health applications closer to everyday clinical and personal use.

4 Empowerment

The third pillar of our research vision focuses on empowering individuals through AI-enabled communication technologies. Speech is fundamental to human interaction, yet effective communication can be compromised in many situations. Patients with speech impairments, such as those who have undergone a laryngectomy, may lose the ability to communicate using their natural voice. In addition, healthy individuals may prefer silent communication to protect privacy in public environments or to communicate effectively in noisy surroundings where audible speech is difficult.

Silent Speech Interfaces (SSIs) offer a promising solution by enabling communication without audible articulation [Schultz *et al.*, 2017]. Instead of relying on acoustic speech, SSI technologies recognise or synthesise speech from alternative biosignals, such as electrocorticography (ECoG), electromagnetic articulography (EMA), and surface electromyography (EMG) [Komeiji *et al.*, 2024; Schultz *et al.*, 2017; Wand and Schultz, 2014]. Among these sensing modalities, our research primarily focuses on EMG because it is non-invasive, wearable, and well suited for practical daily use.

4.1 EMG-based Silent Speech Communication

Benefiting from its non-invasive and wearable nature, surface EMG has become the primary sensing modality in our research on Silent Speech Interfaces. Our work focuses on translating facial muscle activity directly into natural and intelligible speech.

To advance this field, we developed two publicly available EMG speech corpora, EMG-Vox [Scheck *et al.*, 2024] and EMG-VCTK [Scheck *et al.*, 2025], which support research on cross-speaker adaptation, multi-speaker speech synthesis, and zero-shot voice generation. These datasets provide a foundation for developing more generalisable EMG-to-speech systems capable of synthesising speech for both seen and unseen speakers.

Beyond improving intelligibility, our research has focused on enhancing the naturalness and practicality of synthesised speech. We proposed Diff-ETS, which employs diffusion probabilistic models to generate more natural speech and significantly improves subjective speech quality over conventional EMG-to-speech models [Ren *et al.*, 2024b]. To enable real-world deployment, we further developed the Stream-ETS causal architecture, achieving real-time EMG-to-speech conversion with processing latencies as low as 7.72 ms on GPU for 10 ms EMG segments [Scheck *et al.*, 2023]. Together, these advances move EMG-based silent speech communication closer to practical daily use.

4.2 EMG in Silent Paralinguistics

Existing EMG-to-speech research has primarily focused on improving speech intelligibility by recovering linguistic content from biosignals. However, speech also conveys rich paralinguistic information, such as emotion, speaking style, and social intent. To bridge this gap, we introduced the concept of *Silent Paralinguistics* [Ren *et al.*, 2025], which extends Silent Speech Interfaces beyond speech reconstruction to the recognition of paralinguistic information from biosignals and expressive speech synthesis.

As an initial step, we developed multimodal recording protocols incorporating both reading and Wizard-of-Oz interaction tasks, and demonstrated that AI models can reliably recognise affective states, including politeness and frustration, directly from EMG signals [Pistrosch *et al.*, 2026]. This research opens new opportunities for more expressive, natural, and human-centred silent communication systems.

5 Future Insights

The next generation of digital health will be characterised by continuous, personalised, and human-centred healthcare enabled by trustworthy AI. Rather than developing isolated models for individual diseases or sensing modalities, future research should focus on intelligent systems that integrate multimodal biosignals, including speech, body sounds, physiological signals, and wearable sensor data, to provide a holistic understanding of human health. Such systems should be capable of lifelong learning from longitudinal data while adapting to individual users and evolving clinical conditions.

From the perspective of sensing, future AI systems should move beyond analysing a single modality toward multimodal biosignal foundation models that learn transferable representations across diverse healthcare tasks [Zhang *et al.*, 2026]. Combined with advances in wearable sensing technologies, these models could enable continuous monitoring of physical, cognitive, and mental health in everyday environments, facilitating earlier disease detection and more personalised interventions.

From the perspective of understanding, trustworthy AI should extend beyond predictive accuracy to encompass explainability, robustness, fairness, privacy preservation, and uncertainty awareness. Future research should develop AI systems that not only provide accurate predictions but also communicate reliable confidence estimates, explain their reasoning in clinically meaningful ways, and remain robust against distribution shifts, adversarial attacks, and heterogeneous real-world data [Wang *et al.*, 2026]. Integrating these capabilities will be essential for establishing trustworthy AI that can be safely deployed in clinical practice.

From the perspective of empowerment, intelligent healthcare technologies should increasingly function as collaborative partners that assist both clinicians and patients. In addition to restoring communication through silent speech interfaces, future systems should recognise richer physiological and behavioural information, including emotion, stress, fatigue, and cognitive states, enabling more natural human-AI interaction and personalised healthcare support. Ultimately, trustworthy AI [Amadi and Ojo, 2025] should augment rather than replace human expertise, providing timely, transparent, and actionable recommendations while keeping clinicians and patients at the centre of decision making.

Collectively, these research directions point toward a future in which trustworthy AI forms a bridge between multimodal sensing technologies and clinical decision support. By combining reliable sensing, transparent understanding, and human-centred empowerment, AI has the potential to become an intelligent lifelong healthcare companion that improves accessibility, quality of care, and quality of life.

6 Conclusion

This paper has presented a research vision for advancing trustworthy AI in human-centred digital health through three interconnected pillars: Sensing, Understanding, and Empowerment. By enabling machines to perceive clinically relevant information from biosignals, developing AI systems that are transparent, robust, efficient, secure, and privacy-preserving, and empowering individuals through intelligent assistive communication technologies such as EMG-based Silent Speech Interfaces, these research directions collectively address key challenges that hinder the clinical adoption of AI. Looking ahead, advances in multimodal foundation models, wearable sensing technologies, trustworthy machine learning, and human-AI collaboration will further accelerate the integration of AI into everyday healthcare. Ultimately, the goal is to establish trustworthy AI as a bridge between multimodal sensing technologies and clinical decision support, enabling intelligent systems that are reliable, explainable, accessible, and human-centred, thereby improving healthcare delivery, communication, and quality of life.

Ethical Statement

Generative AI tools (ChatGPT, OpenAI) were used solely to assist with language editing and improving the readability of the manuscript. The author reviewed and verified all generated text and takes full responsibility for the accuracy, originality, and integrity of the work.

Acknowledgments

This work is partially funded by the DFG, German Research Foundation, through the project "Silent Paralinguistics" with grant number 514018165.

References

[Amadi and Ojo, 2025] Chimaobi Amadi and Adegboyega Ojo. Building trustworthy AI in healthcare. *IEEE Access*, 14:1182–1212, 2025.

[Basso *et al.*, 2023] Leonie Basso, Zhao Ren, and Wolfgang Nejdl. Towards efficient ECG-based atrial fibrillation detection via parameterised hypercomplex neural networks. In *Proc. EUSIPCO*, pages 1375–1379, Helsinki, Finland, 2023.

[Chambres *et al.*, 2018] Gaëtan Chambres, Pierre Hanna, and Myriam Desainte-Catherine. Automatic detection of patient with respiratory diseases using lung sound analysis. In *CBMI*, 2018. 6 pages.

[Chang *et al.*, 2022] Yi Chang, Zhao Ren, Thanh Tam Nguyen, Wolfgang Nejdl, and Björn W Schuller. Example-based explanations with adversarial attacks for respiratory sound analysis. In *Proc. INTERPSEECH*, pages 4003–4007, Incheon, Korea, 2022.

[Chang *et al.*, 2024] Yi Chang, Zhao Ren, Zixing Zhang, Xin Jing, Kun Qian, Xi Shao, Bin Hu, Tanja Schultz, and Björn W Schuller. STAA-Net: A sparse and transferable adversarial attack for speech emotion recognition. *IEEE Transactions on Affective Computing*, 16(2):861–874, 2024.

[Chang *et al.*, 2025] Yi Chang, Zhao Ren, Zhonghao Zhao, Thanh Tam Nguyen, Kun Qian, Tanja Schultz, and Björn W Schuller. Breaking resource barriers in speech emotion recognition via data distillation. In *Proc. INTERSPEECH*, pages 141–145, Rotterdam, The Netherlands, 2025.

[Chang *et al.*, 2026] Yi Chang, Sofiane Laridi, Zhao Ren, Gregory Palmer, Björn W. Schuller, and Marco Fisichella. A two-stage defence for robust federated speech emotion recognition. In *Proc. INTERSPEECH*, Sydney, Australia, 2026.

[Cummins *et al.*, 2020] Nicholas Cummins, Yilin Pan, Zhao Ren, Julian Fritsch, Venkata Nallanthighal, Heidi Christensen, Daniel Blackburn, Björn W Schuller, Mathew Magimai.-Doss, Helmer Strik, and Aki Härmä. A comparison of acoustic and linguistics methodologies for Alzheimer's dementia recognition. In *Proc. INTERSPEECH*, pages 2182–2186, Shanghai, China, 2020.

[Dong *et al.*, 2019] Fengquan Dong, Kun Qian, Zhao Ren, Alice Baird, Xinjian Li, Zhenyu Dai, Bo Dong, Florian Metze, Yoshiharu Yamamoto, and Björn W Schuller. Machine listening for heart status monitoring: Introducing and benchmarking HSS – the heart sounds shenzhen corpus. *IEEE Journal of Biomedical and Health Informatics*, 24(7):2082–2092, 2019.

[Gerczuk *et al.*, 2023] Maurice Gerczuk, Andreas Triantafyllopoulos, Shahin Amiriparian, Alexander Kathan, Jonathan Bauer, Matthias Berking, and Björn W Schuller. Zero-shot personalization of speech foundation models for depressed mood monitoring. *Patterns*, 4(11):1–15, 2023.

[Jiang *et al.*, 2024] Zifan Jiang, Salman Seyedi, Emily Griner, Ahmed Abbasi, Ali Bahrami Rad, Hyeokhyen Kwon, Robert O Cotes, and Gari D Clifford. Multimodal mental health digital biomarker analysis from remote interviews using facial, vocal, linguistic, and cardiovascular patterns. *IEEE Journal of Biomedical and Health Informatics*, 28(3):1680–1691, 2024.

[Juhn and Liu, 2020] Young Juhn and Hongfang Liu. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy and Clinical Immunology*, 145(2):463–469, 2020.

[Komeiji *et al.*, 2024] Shuji Komeiji, Takumi Mitsuhashi, Yasushi Iimura, Hiroharu Suzuki, Hidenori Sugano, Koichi Shinoda, and Toshihisa Tanaka. Feasibility of decoding covert speech in ECoG with a Transformer trained on overt speech. *Scientific Reports*, 14(1):11491 (1–13), 2024.

[Nasr *et al.*, 2026] Seham Nasr, Zhao Ren, Sudi Murindanyi, and David S. Johnson. TempA: Temporal alignment of human expert-referenced with model-inferred acoustic concepts for speech emotion recognition. In *Proc. IJCAI Workshop EI-MiGA*, Bremen, Germany, 2026.

[Nguyen *et al.*, 2025] Thanh Tam Nguyen, Thanh Trung Huynh, Zhao Ren, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, and Quoc Viet Hung Nguyen. A survey of machine unlearning. *ACM Transactions on Intelligent Systems and Technology*, 16(5):108 (1–46), 2025.

[Peng *et al.*, 2024] Liyizhe Peng, Zixing Zhang, Tao Pang, Jing Han, Huan Zhao, Hao Chen, and Björn W Schuller. Customising general large language models for specialised emotion recognition tasks. In *Proc. ICASSP*, pages 11326–11330, Barcelona, Spain, 2024.

[Pistrosch *et al.*, 2026] Simon Pistrosch, Kleanthis Avramidis, Zhao Ren, Tiantian Feng, Jihwan Lee, Monica Gonzalez-Machorro, Anton Batliner, Tanja Schultz, Shrikanth Narayanan, and Björn W Schuller. Affect decoding in phonated and silent speech production from surface EMG. *arXiv preprint arXiv:2603.11715*, 2026.

[Rammohan *et al.*, 2025] Rathi Adarshi Rammohan, Zhao Ren, Aleksandra Swiderska, Dennis Küster, and Tanja Schultz. Unveiling deep speech embeddings: Acoustic insights into hatespeech detection. In *ITG*, pages 16–20, Berlin, Germany, 2025.

[Ren *et al.*, 2019] Zhao Ren, Jing Han, Nicholas Cummins, Qiuqiang Kong, Mark Plumbley, and Björn W Schuller. Multi-instance learning for bipolar disorder diagnosis using weakly labelled speech data. In *Proc. DPH*, pages 79–83, Marseille, France, 2019.

[Ren *et al.*, 2020a] Zhao Ren, Alice Baird, Jing Han, Zixing Zhang, and Björn W Schuller. Generating and protecting against adversarial attacks for deep speech-based emotion recognition models. In *Proc. ICASSP*, pages 7184–7188, Barcelona, Spain, 2020.

[Ren *et al.*, 2020b] Zhao Ren, Jing Han, Nicholas Cummins, and Björn W Schuller. Enhancing transferability of black-box adversarial attacks via lifelong learning for speech emotion recognition models. In *Proc. INTERSPEECH*, pages 496–500, Shanghai, China, 2020.

[Ren *et al.*, 2022] Zhao Ren, Kun Qian, Fengquan Dong, Zhenyu Dai, Wolfgang Nejdl, Yoshiharu Yamamoto, and Björn W Schuller. Deep attention-based neural networks for explainable heart sound classification. *Machine Learning with Applications*, 9:1–9, 2022.

[Ren *et al.*, 2023a] Zhao Ren, Thanh Tam Nguyen, Yi Chang, and Björn W Schuller. Fast yet effective speech emotion recognition with self-distillation. In *Proc. ICASSP*, pages 1–5, Rhodes, Greece, 2023.

[Ren *et al.*, 2023b] Zhao Ren, Kevin Scheck, and Tanja Schultz. Self-learning and active-learning for electromyography-to-speech Conversion. In *Proc. ITG*, pages 245–249, Achen, Germany, 2023.

[Ren *et al.*, 2024a] Zhao Ren, Yi Chang, Thanh Tam Nguyen, Yang Tan, Kun Qian, and Björn W Schuller. A comprehensive survey on heart sound analysis in the deep learning era. *IEEE Computational Intelligence Magazine*, 19(3):42–57, 2024.

[Ren *et al.*, 2024b] Zhao Ren, Kevin Scheck, Qinhan Hou, Stefano van Gogh, Michael Wand, and Tanja Schultz. Diff-ETS: Learning a diffusion probabilistic model for electromyography-to-speech conversion. In *Proc. EMBC*, pages 1–4, Orlando, FL, 2024.

[Ren *et al.*, 2025] Zhao Ren, Simon Pistrosch, Buket Coşkun, Kevin Scheck, Anton Batliner, Björn W Schuller, and Tanja Schultz. An introduction to silent paralinguistics. *arXiv preprint arXiv:2508.18127*, 2025.

[Ringeval *et al.*, 2019] Fabien Ringeval, Björn W Schuller, Michel Valstar, Nicholas Cummins, Roddy Cowie, Leili Tavabi, Maximilian Schmitt, Sina Alisamir, Shahin Amiriparian, Eva-Maria Messner, Siyang Song, Shuo Liu, Ziping Zhao, Adria Mallol-Ragolta, Zhao Ren, Mohammad Soleymani, and Maja Pantic. AVEC 2019 workshop and challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition. In *Proc. AVEC*, pages 3–12, Nice, France, 2019.

[Scheck *et al.*, 2023] Kevin Scheck, Darius Ivucic, Zhao Ren, and Tanja Schultz. Stream-ETS: Low-latency end-to-end speech synthesis from electromyography signals. In *Proc. ITG*, pages 200–204, Achen, Germany, 2023.

[Scheck *et al.*, 2024] Kevin Scheck, Zhao Ren, Tom Dombeck, Jenny; Sonnert, Stefano van Gogh, Qinhan Hou, Michael Wand, and Tanja Schultz. Cross-speaker training and adaptation for electromyography-to-speech conversion. In *Proc. EMBC*, pages 1–4, Orlando, FL, 2024.

[Scheck *et al.*, 2025] Kevin Scheck, Tom Dombeck, Zhao Ren, Peter Wu, Michael Wand, and Tanja Schultz. DiffMV-ETS: Diffusion-based multi-voice electromyography-to-speech conversion using speaker-independent speech training targets. In *Proc. INTERSPEECH*, pages 5573–5577, Rotterdam, The Netherlands, 2025.

[Schuller and Batliner, 2013] Björn Schuller and Anton Batliner. *Computational paralinguistics: emotion, affect and personality in speech and language processing*. John Wiley & Sons, 2013.

[Schuller *et al.*, 2018] Björn W Schuller, Stefan Steidl, Anton Batliner, Peter Marschik, Harald Baumeister, Fengquan Dong, Simone Hantke, Florian Pokorny, Eva-Maria Rathner, Katrin Bartl-Pokorny, Christa Einspieler, Dajie Zhang, Alice Baird, Shahin Amiriparian, Kun Qian, Zhao Ren, Maximilian Schmitt, Panagiotis Tzirakis, and Stefanos Zafeiriou. The INTERSPEECH 2018 computational paralinguistics challenge: Atypical & self-assessed affect, crying & heart beats. In *Proc. INTERSPEECH*, pages 122–126, Hyderbad, India, 2018.

[Schultz *et al.*, 2017] Tanja Schultz, Michael Wand, Thomas Hueber, Dean J. Krusienski, Christian Herff, and Jonathan S. Brumberg. Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 25:2257–2271, 2017.

[Wand and Schultz, 2014] Michael Wand and Tanja Schultz. Towards real-life application of EMG-based speech recognition by using unsupervised adaptation. In *Proc. INTERSPEECH*, pages 1189–1193, Singapore, 2014.

[Wang *et al.*, 2026] Yishan Wang, Tsai-Ning Wang, Mathias Funk, and Aaqib Saeed. StethoLM: Audio language model for cardiopulmonary analysis across clinical tasks. *Transactions on Machine Learning Research*, 2026.

[Xu *et al.*, 2026] Yiwen Xu, Tariq M Khan, Yang Song, and Erik Meijering. Edge deep learning in computer vision and medical diagnostics: A comprehensive survey. *arXiv preprint arXiv:2605.06714*, 2026.

[Yeung *et al.*, 2023] Andy Wai Kan Yeung, Ali Torkamani, Atul J Butte, Benjamin S Glicksberg, Björn Schuller, Blanca Rodriguez, Daniel SW Ting, David Bates, Eva Schaden, Hanchuan Peng, et al. The promise of digital healthcare technologies. *Frontiers in public health*, 11:1196596, 2023.

[Zhang *et al.*, 2026] Yuwei Zhang, Kumar Ayush, Siyuan Qiao, A. Ali Heydari, Girish Narayanswamy, Max Xu, et al. SensorLM: Learning the language of wearable sensors. In *Proc NeurIPS*, Sydney, Australia, 2026.