Embedded AI and Sensing for Wellness, Health, and Sustainability: Intelligent, Privacy-Aware Wearable Devices and AIoT Systems

Jingping Nie

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

© 2025

Jingping Nie

All Rights Reserved

Abstract

Embedded AI and Sensing for Wellness, Health, and Sustainability:

Intelligent, Privacy-Aware Wearable Devices and AIoT Systems

Jingping Nie

With the rise of connected devices and the Artificial Intelligence of Things (AIoT), there is immense potential to transform personal wellness, health care, and environmental responsibility through embedded AI, sensing, and machine learning technologies. However, existing solutions often fall short in terms of accessibility, privacy, and sustainability, motivating the need for innovative, human-centric systems that can seamlessly blend into daily life while maximizing social welfare and sustainability. In this thesis, we explore the design and deployment of intelligent, privacy-aware, and pervasive wearable devices and AIoT systems to enhance personal health, well-being, and sustainability.

Motivated by these challenges, my broader research spans four interconnected thrusts: (*1*) smart wearables and platforms for emotion, fitness, and wellness; (*2*) model-driven biosignal analysis and emotion detection; (*3*) AI-based mental and physical health care in smart home environments; and (*4*) city-scale human-in-the-loop EV-interfaced grid optimization. This thesis presents the major topics and representative systems developed across these four thrusts. Unifying these efforts is a shared human-centric design philosophy that begins with sensing the individual and extends to personalized support from the individual level to city scale.

The first topic introduces low-cost wearable systems for real-time monitoring of emotion and fitness, including wireless glasses for facial expression classification and real dimensional emotion

monitoring. The second topic explores contactless audio-based platforms for treadmill running metrics estimation and model-driven analysis of physiological signals using machine learning and acoustic foundation models on post-exercise speech and biosignals.

In the third topic, we design AI-enhanced smart home solutions for daily functioning and mental health support. Through collaborations with psychotherapists, we develop an intelligent home-based screening and intervention framework that integrates ambient sensing, conversational agents, and robotic feedback for scalable, in-situ mental health and day-to-day functioning support.

Scaling to city scale, this thesis, in the fourth topic, advances human-centric optimization of electric vehicle (EV)-interfaced microgrids to support sustainable urban energy systems. We first develop a constrained optimization solver and a reinforcement learning–based control framework to estimate and regulate power flows under partial observability and IEEE-standardized grid services. Building on this, we design a practical, data-driven EV charging recommendation system that accounts for user preferences, charger availability, pricing, and grid capacity. Together, these tools enhance the integration of EVs into smart grids while co-optimizing outcomes for drivers, charger owners, and utility operators.

Across all topics, this dissertation emphasizes hardware-software co-design, real-world deployment, and multidisciplinary collaboration. It proposes a unified framework that senses, interprets, and learns from human behavior through multimodal sensing data, enabling intelligent systems to deliver personalized recommendations and interventions for fitness, mental and physical health, sustainability, and overall well-being. Through rigorous evaluation, public datasets, and open-source tools, this work demonstrates how embedded AI and AIoT systems can meaningfully enhance both individual wellness and community-scale services. The findings point toward a future where pervasive, privacy-aware technologies are seamlessly integrated into daily life, supporting proactive care, sustainable living, and human-centered design at scale.

Table of Contents

Acknowledgments xxii

Dedication xxv

Chapter 1: Introduction 1

1.1 Main Topics and Contributions 3

1.1.1 Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition 4

1.1.2 Contactless Acoustic-based System for Real-time Running Metrics Monitoring and Post-Exercise Body Signal Analysis 5

1.1.3 AI-based Mental and Physical Health Care in Smart Home Environments . 7

1.1.4 Human-Centric Data-Driven Optimization and Recommendation in EV-Interfaced Grid at City Scale 7

1.2 Contributions to Literature 9

Chapter 2: Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition 11

2.1 Related Work 14

2.1.1 Bio-signal Acquisition by Wearable Devices 14

2.1.2 Vision-Based Apparent Emotion Classification 15

2.1.3 Real Emotion Recognition 16

2.2 SPIDERS+System Design 16

2.2.1 Sensing Modalities Overview 18
2.2.2 Hardware Platform 19
2.2.3 Price Breakdown 22
2.2.4 Power Consumption 23
2.2.5 Heat Dissipation and Comfort 23
2.2.6 System Pipeline and Typical Usage 24
2.3 Core Function Library 25
2.3.1 Eye & Eyebrow Shape Detection 26
2.3.2 Pupillometry 29
2.3.3 Zygomaticus Muscle Movements 32
2.3.4 Heart Rate 33
2.4 Advanced Functionalities Prototyping 35
2.4.1 Facial Expression Detection 36
2.4.2 Real Emotion Classification 41
2.5 Enabled Applications 43
2.5.1 Fatigue Detection 44
2.5.2 Safe Driving and Concentration Monitoring 45
2.5.3 Mental Health Monitoring 46
2.5.4 Entertainment and Gaming 47
2.5.5 Marketing and Advertisements 47
2.5.6 Elderly Care 48
2.6 Desktop and Smartphone Applications 48

Chapter 3: Contactless Acoustic-based System for Real-time Running Metrics Monitoring and Post-Exercise Body Signal Analysis 50

3.1 Background and Related Work 55

3.1.1 Running Metrics Background 55

3.1.2 Running-Form-based Running Metrics Estimation Methods 56

3.1.3 Treadmill Running Dataset 58

3.1.4 Cardiorespiratory Audio Datasets 59

3.1.5 Speech Breathing and Disfluency and Downstream Applications 59

3.2 Data Collection 60

3.2.1 Data Collection Procedures 61

3.2.2 Data Synchronization 63

3.3 SoundTrackDB Running Metrics Annotation Generation and Validation 64

3.3.1 YOLO-based Ground-truth Generation Method 64

3.3.2 CV-based Ground-truth Generation Evaluation 65

3.4 Description of SoundTrackDB and Post-Run Datasets 67

3.4.1 SoundTrackDB Overview 67

3.4.2 Post-Run Body Sound Dataset Overview 69

3.5 SoundTrack System Architecture 70

3.6 SoundTrack Training and Microbenchmarks 71

3.6.1 Denoising Methods 72

3.6.2 Features 73

3.6.3 Models and Training 74

3.6.4 Microbenchmarks 76

3.6.5 SoundTrack Evaluation 81

3.7 Post-Run Body Sound Analysis Training and Microbenchmarks 83

3.7.1 Features, Labels, and Model Training 83

3.7.2 Exertion Level Classi cation Evaluation 84

3.8 SoundTrack Implementation and Case Study 85

3.8.1 SoundTrack Implementation 85

3.8.2 SoundTrack Case Study in the Wild 86

Chapter 4: AI-Enhanced Smart Home Solutions for Accessible Daily Functioning and Mental Health Screening & Psychotherapies 91

4.1 Psychological Background 96

4.1.1 Psychological Assessment 96

4.1.2 Psychotherapeutic Interventions 96

4.2 Related Work 98

4.2.1 Mental Wellness Self-Screening and Self-Care 98

4.2.2 LLM-based Healthcare and Mental Healthcare 99

4.2.3 Sensor-based Activity Recognition 100

4.2.4 2D LiDAR-based Indoor Human Detection and Tracking 101

4.2.5 Camera-based Object Detection and Posture Detection 102

4.3 CaiTI System Architecture 103

4.3.1 Activity Detection in Ambient Sensing Home Environment 105

4.3.2 Conversation Principles and Underlying Rationale 105

4.3.3 Conversation Generation, Analysis, and Psychotherapeutic Intervention . . 107

4.3.4 Home Robot Check-in 109

4.4 Method for M1 and M4 112

4.4.1 Sensor-based Activity Detection in Ambient Sensing Home Environment 112

4.4.2 Interaction and Precautionary Check-in with Home Robot 115

4.5 Method for M2 and M3 and Microbenchmarks 119

4.5.1 CaiTI's Questioner 120

4.5.2 Response Analyzer 121

4.5.3 Re ection-Validation Reasoner , Guide , and Validator 122

4.5.4 Cognitive Behavioral Therapy Reasoner and Guide 126

4.5.5 Discussion and Observations on different LLMs 129

4.6 Implementation and Study Design 129

4.6.1 Subject Recruitment 130

4.6.2 Prototyping Hardware Deployment 130

4.6.3 Subject Recruitment 130

4.6.4 LLMs Implementation 130

4.6.5 Implementation 131

4.6.6 Study Design 133

4.7 CaiTI System Evaluation 135

4.7.1 M1: Activity Detection in Ambient Sensing Home Environment 135

4.7.2 M2: Converse with the User in a Natural Way & M3: Conversational Intervention 136

4.7.3 M4: Precautionary Check-in with Home Robot 139

4.8 User Study 142

4.8.1 User Adaptation Indicated by Device Usage 142

4.8.2 Quantitative Analysis 142

4.8.3 Qualitative Evaluation from the Subjects and Therapists 143

4.9 Discussion 146

Chapter 5: Optimal Power Flow Estimation and Recommendation for EV-Interfaced Microgrids with Standardized Grid Services 148

5.1 Background 153

5.1.1 Grid Structure 153

5.1.2 Distributed Energy Resources and Grid Services 154

5.1.3 Grid Service Control Mechanism for Renewable Energy-Interfaced Microgrid 155

5.1.4 AC Power Flows 157

5.2 Estimation based on AC Power Flows with Constrained Optimization Solver 158

5.2.1 Grid State Estimation Based on AC Power Flows Method 158

5.2.2 Estimation Accuracy Evaluation on IEEE 30-Bus System 159

5.2.3 Performance Comparison with pandapower 160

5.2.4 Comparison to MATPOWER State Estimation Function 163

5.2.5 Estimation Accuracy on Various IEEE Bus Systems 164

5.3 Exploratory Study of Grid Services Implementation with Constrained Optimization Solver 164

5.3.1 Grid Services Implementation 166

5.3.2 Grid Services on IEEE-9 bus 166

5.3.3 Grid Services on IEEE 57-Bus and 118-Bus System 170

5.3.4 Sensitivity Analysis on IEEE 30-Bus System 171

5.4 Deep Reinforcement Learning Based System Pipeline 173

5.4.1 System Pipeline 174

5.4.2 Training Reward and Grid Service Constraints 176

5.5 Exploratory Study of Grid Services Implementation with Deep Reinforcement Learning 177

5.5.1 Experimental Setup 177

5.5.2 Training Results and Evaluation 177

Chapter 6: Human-Centric Data-Driven Optimization and Recommendation in EV-Interfaced Grid at City Scale 181

6.1 Related Work 184

6.1.1 EV Charger & Charging Session Information 184

6.1.2 EV SoC Infomation 185

6.1.3 Vehicle Mobility 185

6.1.4 Grid Capacity 186

6.1.5 EV Charging Optimization 187

6.2 Recommender System Design 188

6.3 Vehicle Data Preparation 192

6.3.1 City-Scale Mobility Data and Augmentation 192

6.3.2 Driving SoC Data 193

6.3.3 SoC Data Augmentation 193

6.3.4 EV Driving Dataset for the DQN-based DRL Recommender 198

6.4 EV Charger Data Preparation 198

6.5 Deep Reinforcement Learning Architecuture 200

6.5.1 Practical Modeling of the Reward Function 202

6.5.2 Driving Behaviour Information Remapping 204

6.6 Evaluation 205

6.7 Limitation and Discussion 208

Conclusion or Epilogue 210

References 213

Appendix A: Dimensions for Day-to-day Functioning Screening. 251

List of Figures

1.1 My research focuses on four interconnected thrusts spanning from the individual level to the system level: (Thrust 1) smart wearables and platforms for emotion, tness, and wellness [1, 2, 3, 4, 5, 6, 7, 8]; (Thrust 2) model-driven biosignal analysis and emotion detection [9, 10, 11, 12, 13]; (Thrust 3) AI-based mental and physical health care in smart home environments [14, 15, 16, 17, 18, 19, 20, 21]; and (Thrust 4) city-scale human-in-the-loop electric vehicle (EV)-interfaced grid optimization [22, 23, 24, 25, 26]. 2

1.2 The four topics of this thesis. 4

2.1 Left: The SPIDERS+ glasses platform including hardware sensors (red), PCB boards (blue), frame components (yellow), and optional attachments (green). Right: SPIDERS+ glasses on a mannequin. 12

2.2 The four-layer system architecture of SPIDERS+. We designed the architecture in a top-down approach based on the types of applications that SPIDERS+ aims to impact. All components and layers in SPIDERS+ work towards enabling or improving the top layer – Enabled Applications 17

2.3 PCB layout of the mainboard (right) and biopotential measurement board (left) . . 21

2.4 Thermal image of SPIDERS+ in the active mode. 21

2.5 The ow diagram of the functions in SPIDERS+ client application and the user-facing smartphone/desktop application. 24

2.6 Breakdown of our proposed CNN for eye shape and eyebrow landmark generation. 26

2.7 Left: Example raw image of the eye. Right: Eye image with 27 landmark points generated by the CNN (9 points on the upper eyelid, 9 points on the lower eyelid, and 9 points on the eyebrow). 27

2.8 Training and validation loss during the training process of the landmark generation CNN, where the validation loss (red line) converges in around 300 epochs. 27

2.9 Comparison between the ground-truth landmarks and the landmarks output by the CNN model for regular clear images captured by SPIDERS+ (red: hand-labeled ground truth; green: output of the CNN model). 28

2.10 Comparison between the ground-truth landmarks and the landmarks output by the CNN model for extreme cases with blurry images or with squint eye (red: hand-labeled ground truth; green: output of the CNN model). 28

2.11 Pupil data processing pipeline. (a): The original image captured by SPIDERS+. (b): The 27 landmarks generated by the CNN model mentioned in Section 2.3.1. (c): The cropped eye area image is based on the landmarks for upper and lower eyelids. (d): The color histogram of (c), which is used to detect the threshold of the pupil light intensity. (e): Binary image based on the threshold (white: pupil). (f): The raw contour of the pupil. (g): The raw contour curvature and its segmentation according to the negative peaks. (h): The contour of the pupil. (i): The elliptic contour of the pupil. 30

2.12 A 60-second segment of pupil diameters measured by both Tobii Pro Nano and SPIDERS+. 33

2.13 The relative distances between the proximity sensor and the zygomaticus muscle while a subject alternates smile and calm facial expressions for 10 repetitions among 5 different trials (red shades: the ground truth when the subject is asked to smile). 33

2.14 The ltered and smoothed PPG (blue solid line), estimated heart beat peaks (pink circles), calculated heart rate in BPM (orange stars), and the ground truth provided by a commercial heart rate monitor (blue dashed line). 34

2.15 Eye images of seven common facial expressions. The constraints from the gray-scale image captured by a wearable eye tracker make it dif cult to distinguish subtle differences among different emotions (e.g., happy and disgusted). 35

2.16 The pipeline of eye-image-based facial expression (apparent emotion) detection. . . 38

2.17 Example of binary thresholding conducted on baseline corrected eye size data (green shades: the ground truth when the subject is asked to smile to express happy). 38

2.18 Motion tracking between two consecutive image frames (top). As shown in the motion vector map (bottom), the eyebrow region moves primarily leftwards from the image on the left to the image on the right. 39

2.19 Tracking of eyebrow region motion in the form of overall motion vector components over time (green shades: the ground truth when the subject is asked to perform disgusted). 39

2.20 Left: Eye row motion distribution. Right: Eye size histogram for the same subject. . 40

2.21 Confusion matrix for facial expression classi cation. 42

2.22 GUI of the application used to collect data for training and evaluating our real emotion classi er. Users are shown a stimulus through the GUI and are asked to select their arousal and valence level after each stimulus. 42

2.23 Confusion matrix for arousal detection. 43

2.24 Confusion matrix for valence detection. 43

2.25 The processing pipeline for fatigue detection. 44

2.26 The interface of the user panel on the web-page-based desktop application. 49

2.27 Screenshots of SPIDERS+'s phone application. Top: The homepage of the smartphone application. Bottom: The user panel with the same layout as the desktop application. The alarm noti cation pops up along with a warning sound, if turns on the alarm. 49

3.1 SoundTrack: A contactless solution that uses mobile platforms to accurately estimate cadence and ground contact time (GCT) during treadmill running with on-device machine learning, suitable for runners of all levels and compatible with various treadmill models in gyms with different environmental noise, without needing user-speci c calibration. It provides real-time visualizations of running metrics and personalized feedback and suggestions. 52

3.2 The data collection process is divided into three phases: Pre-Run, In-the-Run, and Post-Run sessions, capturing treadmill running acoustics (audio and IMU), audio data. 61

3.3 Post-Run speech, breathing, and phonocardiogram collection procedure. 62

3.4 Data syncing, processing, and human annotation procedure to generate ground-truth running metrics through CV method for SoundTrackDB. 64

3.5 Top two rows: Performance comparison of SoundTrackDB, RunScribe, and a sports watch for cadence estimation relative to manual annotations. Bottom row: Distribution of cadence estimation errors for SoundTrackDB, the MediaPipe-based method from [5], and RunScribe. 65

3.6 Top two rows: Comparison of GCT estimation performance among SoundTrackDB and RunScribe, benchmarked against manual annotations. Bottom row: Distribution of GCT estimation errors for SoundTrackDB, the MediaPipe-based method from [5], and RunScribe. 66

3.7 Overview of SoundTrackDB to be open sourced: treadmill running sound, CV-based, sports watch and RunScribe collected running metrics, background information, and user study. 67

3.8 The example average decibel (dB) levels for the environment (treadmill off), treadmill operating (without people running), and treadmill running sound (a runner running on the operating treadmill) in different locations from randomly selected sessions. 70

3.9 Overview of the Post-Run Dataset. 70

3.10 SoundTrack system architecture. 71

3.11 SoundTrack mobile App demonstration, which contains: (1) Dashboard Page that displays summary metrics; (2) Set Target Page that allows the user to set specific targets for cadence and GCT as well as choose the visualization modes, focus or trend, during the run; (3) Focus Mode Page that provides real-time running metrics and suggestion; and (4) Trend Mode Page that offers a trend analysis of cadence and GCT during the run in addition to the real-time metrics. 72

3.12 Three prediction methods experimented: (1) Seq2Seq Detection, (2) Classification, and (3) Regression 73

3.13 Three types of models trained for each model architecture: (1) single-task GCT, (2) single-task cadence, and (3) multi-task model leveraging shared acoustic attributes for GCT and cadence estimation. 75

3.14 (a): the effect of various denoising methods on the MFCC of an example 3-second window, with the MVN & Wiener Filtering approach yielding the most effective features for GCT-related sounds (boxed in green). (b) and (c): Comparison of error distribution for Cadence and GCT after different denoising methods for the testing set. 77

3.15 Neural network structure of regression-based multi-task MLP model with MFCC+RMS as input features. 80

3.16 Session-wise error distribution comparison for Cadence and GCT estimation in Test Set. 81

3.17 Comparison of Cadence and GCT Errors per window across speed, location, and subject in test set. (Note that sport watches do not provide GCT for treadmill running.) 82

3.18 Number of sessions and number of clips for structured paragraph reading and spontaneous speech. 83

3.19 Confusion matrix combining the results of 5-fold cross-validation for the 1D CNN-LSTM model using W2V2 Emb-4 features on spontaneous speech data. 85

3.20 Case study session-wise error distribution comparison for Cadence and GCT estimation. 87

3.21 Error and distribution comparison of Cadence and GCT for running sessions with/without incline in Case Study III. 89

3.22 Quantitative scoring of SoundTrack by the 11 new subjects in the case study in the areas of (i) Usability, (ii) Performance, (iii) Privacy, and (iv) Future Use. 90

4.1 CaiTI is a seamlessly connected home "AI therapist" that screens and assesses the day-to-day functioning of the occupant across 37 dimensions and provides appropriate psychotherapeutic interventions and empathic validations depending on the physical and mental status of the occupant. CaiTI incorporates four mix-and-match modules: M1: Activity Detection in Ambient Sensing Home Environment; M2: Converse with the User in a Natural Way for day-to-day functioning screening; M3: Psychotherapeutic Conversational Intervention; and M4: Precautionary Check-in with Home Robot. 93

4.2 The system architecture of CaiTI consists of two main functionalities: 1) day-to-day functioning screening and 2) precautionary interventions with four modules (M1: green, M2: blue, M3: yellow, and M4: red). Note that M1 and M4 are optional, and users can select based on their preferences and needs. 103

4.3 The work ow of CaiTI, including the connections between the main components and the locations where different functionalities take place. 104

4.4 The process to converse with the user naturally to screen day-to-day functioning (M2) and provide Psychotherapeutic Conversational Intervention (M3, boxed in red) during the conversation. To simulate how therapists usually perform, MI therapy is conducted throughout the conversation, while CBT is only implemented at the end of the conversation session with one problematic dimension that the user wants to work on more. 106

4.5 The example process of how CaiTI follows the sequence based on the Q-table, segments the user response, classi es the segment into the format of (Dimension, Score), and proceeds with re ection-validation (R-V) process. 108

4.6 An example conversation using CaiTI's customized smartphone platform. The MI and CBT embedded in the conversation are annotated. 110

4.7 The user interface where the user can manually select the dimensions to work on. . 111

4.8 Flow diagram for the Precautionary Check-in with Home Robot (M14) process. Being privacy-aware, CaiTI only turns on the privacy-invasive camera when it localizes the human using the low- delity 2D LiDAR. 112

4.9 Activities detected by CaiTI in the ambient sensing home environment, their corresponding sensors, and sensor placement. The activities needing Precautionary Interventions are boxed in red. 112

4.10 Characteristics of the shower and drink events, where shaded area denotes the time duration of the corresponding event. 113

4.11 The characteristics of cook and smoke events (shaded area denotes when the corresponding event occurs). 114

4.12 Characteristics of laundry and exercise events (shaded area denotes when the corresponding event occurs). 114

4.13 Normalized IMU data and sleep event features during one example night for a subject (shaded area denotes when the corresponding event occurs). This subject slept less restfully at the beginning and in the end, as well as between 1 to 2 a.m. . . 115

4.14 Architecture for human following using 2D LiDAR. 115

4.15 The process to localize the legs of the user within the room map (the two legs of the user are circled in red, and other noisy laser point clusters are circled in blue). . 116

4.16 VLM (LLaVA-1.5) with LLM reasoner (GPT-4) method. 119

4.17 The re ection-validation (R-V) pipeline illustrates how the R-V Reasoner is utilized for reasoning when the user provides a valid follow-up response that necessitates further attention, including two example scenarios: Scenario 1, where the user provides a valid follow-up response on the rst attempt; and Scenario 2, where the user initially fails to provide a valid follow-up response, and CaiTI guides the user towards providing valid follow-up responses. 124

4.18 The pipeline of the CBT process, which includes Questioner s, Reasoner s and Guide s. The CBT process in CaiTI contains three stages: CBT_Stage1 – Recognize Negative Thoughts, CBT_Stage2 – Challenge Negative Thoughts, and CBT_Stage3 – Reframe Thoughts. The example scenarios where the user properly undergoes and fails to undergo each CBT stage are presented. 127

4.19 The LLM implementation for different tasks during the study. 131

4.20 Plug-and-play CaiTI sensor platform for contactless activity detection in the home environment. 132

4.21 The home robot is based on iRobot Roomba 690 Robot with a Raspberry Pi 4B, a Pi camera, and a 2D LiDAR connected. 133

4.22 The modules and tasks involved in the 14-day study and 24-week longitudinal study. 133

4.23 (a) The setup for the in-lab session with the sensor platforms labeled by the red boxes. (b) and (c) A subject chats with CaiTI via the smartphone/Amazon Echo at the end of an in-lab session. 134

4.24 (a) Example sensor platform placement to detect cooking and clothes-washing activities with sensor platforms securely placed next to the stove or on the washing machine. (b) Home page of CaiTI's smartphone platform. (c) CaiTI's smartphone interface allows users to log activity start and end times, which serve as ground truth for performance evaluation. (d) A subject chats with CaiTI on her computer at home. 135

4.25 Daily activity detection accuracy. (type I Error: detect a false event. type II Error: fail to detect an event.) 136

4.26 CaiTI's performance in assigning Score to segments of user responses. 136

4.27 The evaluation on the tracking capability of CaiTI. 140

4.28 The performance of the Human Detection and Activity Detection functionalities in CaiTI. 141

4.29 Posture detection accuracy. 141

4.30 The user adaptation to sensor platform (M1), conversation sessions (M2 & M3), and home robot (M4). 142

4.31 The average ratings from 20 subjects on the first day (Day 1) and the last day (Day 14) of the study. The questions shaded in orange, blue, green, and red are related to the overall system, M2 & M3, M1, and M4, respectively. 144

5.1 The network topology of IEEE 9-bus, 30-bus, 57-bus, and 118-bus system. 149

5.2 Volt-Var curve. 152

5.3 Watt-Var curve. 154

5.4 Grid services control diagram for a microgrid interfaced with renewable energy. . . 156

5.5 The MAE for active power, power factor, reactive power, and voltage under different numbers of unknown measurements for 30-bus system using the proposed solver, State Estimation in pandapower, and run_se.m in MATPOWER. For each number of unknown variable (−), dot and error bar represent the MAE and its standard deviation of absolute errors. 161

5.6 Comparison between the estimate values and expected values for 9 buses without grid services implemented. (Red dots: Expected values in steady-state; Blue dots: Values calculated by the solver in steady-state; Bus numbers circled in red: known variables.) 165

5.7 Comparison between the estimate values and expected values for 9 buses with grid services implemented on the unknown variables of q (bus 2, 5, 6, and 9) in Volt-Var mode. In (d): black solid line: $E = V_{Ref} = 1$; black dashed line: $E = V_{Ref}$ $0.02V_{\#}$; and gray dashed line: $E = V_{Ref}$ $0.08V_{\#}$. (Red dots: Expected values in steady-state; Blue and green dots: Values calculated by the solver in under Volt-Var constraints with $(_{\#} = 1$ and $(_{\#} = 5$, respectively; Bus numbers circled in red: known variables.) 167

5.8 (a) shows the absolute distances between the calculated reactive power and the AC steady-state values $(_{\#}$ under each $(_{\#}$, while (b) shows the distances from the voltages provided by the solver to the reference voltage $E = V_{(Ref}$ under each $(_{\#}$. (Blue, pink, orange, purple, and green dots: Values calculated by the solver in under Volt-Var constraints with $(_{\#} = 1, 2, 3, 4$, and 5, respectively; Red dots: Expected values in steady-state; Red bus numbers: known variables.) 168

5.9 Comparison between the estimate values and expected values for 9 buses with grid services implemented on the unknown variables of q (bus 2, 5, 6, and 9) in Watt-Var mode. In (a): black solid line: $? = 0$; black dashed lines: $? = 0.5\%^{0}_{rated}$ & $? = 0.5\%^{0}_{rated}$; and gray dashed lines: $? = \%^{0}_{rated}$ & $? = \%^{0}_{rated}$. (Red dots: Expected values in steady-state; Blue and green dots: Values calculated by the solver in under Watt-Var constraints with $(_{\#} = 1$ and $(_{\#} = 5$, respectively; Red bus numbers: known variables.) 169

5.10 Under different $(_{\#}$, the absolute differences between the calculated reactive power and the AC steady-state values (Q_i) are shown in (a). The distances between the active power values provided by the solver and the reference active power (P_i^0) are illustrated in (b). (Blue, pink, orange, purple, and green dots: Values calculated by the solver in under Watt-Var constraints with $(_{\#} = 1$–2–3–4, and 5, respectively; Red dots: Expected values in steady-state; Bus numbers circled in red: known variables.) 171

5.11 Comparison between the estimate values and expected values for IEEE 57-bus system with grid services implemented in Volt-Var mode. In (d): black solid line: $E = V_{Ref} = 1$; black dashed lines: $E = V_{Ref} \quad 0{\bullet}02V_{\#}$; and gray dashed lines: $E = V_{Ref} \quad 0{\bullet}08V_{\#}$. (Red dots: Expected values in steady-state; Blue and green dots: Values calculated by the solver in under Volt-Var constraints with $(_{\#} = 1$ and $(_{\#} = 5$, respectively.) 172

5.12 Comparison between the estimate values and expected values for IEEE 118-bus system with grid services implemented in Watt-Var mode. In (a): black solid line: $? = 0$; black dashed lines: $? = 0{\bullet}5\%_{rated}$ & $? = 0{\bullet}5\%^{0}_{rated}$; and gray dashed lines: $? = \%_{rated}$ & $? = \%^{0}_{rated}$. In (d): black solid line: $E = V_{Ref} = 1$. (Red dots: Expected values in steady-state; Blue and green dots: Values calculated by the solver in under Watt-Var constraints with $(_{\#} = 1$ and $(_{\#} = 5$, respectively; Red bus numbers: known variables.) 172

5.13 The sensitivity analysis for Volt-Var mode for IEEE 30-bus system (black solid line: $E = V_{Ref} = 1$; black dashed lines: $E = V_{Ref} \quad 0{\bullet}02V_{\#}$; and gray dashed lines: $E = V_{Ref} \quad 0{\bullet}08V_{\#}$.) 173

5.14 The sensitivity analysis for Watt-Var mode for IEEE 30-bus system (black solid line: $? = 0$; black dashed lines: $? = 0{\bullet}5\%_{rated}$ & $? = 0{\bullet}5\%^{0}_{rated}$; and gray dashed lines: $? = \%_{rated}$ & $? = \%^{0}_{rated}$). 174

5.15 System pipeline for generating grid services policies by optimizing the microgrid using a DRL-based approach. (The blue arrows indicate the direction of parameter updates for the neural networks, and the red star represents the place to add noise during the training process to prevent over tting.) 175

5.16 The average training reward for all steps in each episode using TD3, DDPG, and DQN for Constant-Var, Volt-Var, and Watt-Var modes. (Green line: Constant-Var, Blue line: Watt-Var, Red line: Volt-Var.) 178

5.17 The voltage optimization results on each bus for Constant-Var, Volt-Var, and Watt-Var modes using TD3, DDPG, and DQN. (Black dots: original voltage before optimization; Green dots: Constant-Var, Blue dots: Watt-Var, Red dots: Volt-Var.) 180

5.18 The resulting reactive power of each bus after being optimized by TD3, DDPG, and DQN for Constant-Var, Volt-Var, and Watt-Var modes. (Black dots: original voltage before optimization; Green dots: Constant-Var; Blue dots: Watt-Var; Red dots: Volt-Var.) 180

6.1 Data preparation and augmentation ow diagram of the proposed DQN-based DRL recommendation system. 189

6.2 (Left) the process of making recommendations to EV drivers at each timestep, and (right) example screenshots of a customized smartphone app designed for the users using our proposed DRL-based recommendation system. 191

6.3 Location map of the chargers in Manhattan, NYC, that are selected as the agents in the proposed DRL framework. Green markers denote the chargers that are, on average, furthest away from the 61–392 vehicle mobility traces from our augmented dataset and black markers denote the chargers that are, on average, closest to these traces. 199

6.4 The proposed Deep-Q-Network- (DQN-) based deep reinforcement learning (DRL) architecture. 200

6.5 A demonstration of the driver location (lat, lon) and SoC remapping after selected by an agent (charger). In this example, the EV driver 3 will drive towards the charger (agent) 2 after it is selected, and will return back to its original route and continue the trip after it is released by charger 2 with $G_{34C>DA}$ = 7 min and $G_{2\ 0A68=6}$ = 40 min. 204

6.6 DRL training reward across the episodes. 206

6.7 The average driver-to-charger distance (km) in one day for all driver-charger pairs generated by the DRL model trained after 10 (blue) vs. 2,000 (orange) episodes. The shadowed area indicates the improvements that the model made from episodes 10 to 2,000. 207

6.8 The number of drivers on the road during the day and the number of times that the 160 chargers switch to select a different EV driver. 207

6.9 The difference in grid capacity information caused by different charger selections using the model saved in episode 10 (left) and episode 2,000 (right) at an example timestamp. Blue markers denote car location at that timestamp, black markers are all EV chargers and red markers denote chargers selected by the agent. 208

List of Tables

2.1 Price Breakdown and Power Consumption 22

2.2 Expressions and eye area features. 25

2.3 Expressions and eye area features 36

3.1 Exertion level de nition of the modi ed ve-level Borg RPE. 62

3.2 Average MAE Across All Sessions for Cadence and GCT Estimation at Different Speeds Using Various Approaches 65

3.3 Micro-benchmarks of model estimation performances selected from the 9-Fold cross-validation, reporting the accuracy achieved by each type of model (single-task & multi-task) on the test set, the features they used, and power consumption on a smartphone (iPhone 15 Pro Max)Note that GCT in SoundTrackDB are derived from 60 fps video, each missed frame equates to a 16 ms error. 78

3.4 Micro-benchmarks of custom feature processing time implemented on a smartphone (iPhone 15 Pro Max). 79

3.5 Micro-benchmarks of TFLite model inference time on a smartphone (iPhone 15 Pro Max). 79

3.6 Micro-benchmarks for the feature selection process of the multi-task MLP model. 11 acoustic feature combinations were tested, with MFCC+RMS producing the best result. 79

3.7 9-Fold Cross-Validation results of the Multitask-MLP model with MFCC+RMS features. 80

3.8 5-Fold CV results of the combinations of model structures, features, and reading/spontaneous speech. 84

3.9 Results comparing SoundTrack's performance on SoundTrackDB, 26 regular sessions in Case Study I, 2 varying-speed fartlek running sessions with SoundTrack implemented on NordicTrack 2450 treadmill in Case Study II, and additional tests in Case Study III. 87

4.1 Number of bounding boxes in the dataset. 117

4.2 Performance on the testing set for privacy-aware camera-based object detection and activity recognition. 118

4.3 Performance in classifying concerning smoking- and drinking-alone behavior. . . . 119

4.4 Score and Dimension assignment performance comparisons for Response Analyzer using different LLM models. 122

4.5 The performance for R-V Reasoner (0: Valid, 1: Invalid), R-V Guide, and R-V Validator with different LLM models. 126

4.6 CBT Reasoner performance for using different LLM models (0: Valid, 1: Invalid). 128

4.7 CBT Guide accuracies for using different LLM models. 129

4.8 Number of times that CBT Reasoner and CBT Guides triggered in each attempt and the number of CBT sessions terminated in the three CBT stages. 137

4.9 The performance for R-V Reasoner (0: Valid, 1: Invalid), R-V Guide, and R-V Validator in handling the responses from subjects during the experiment. 138

4.10 CBT Reasoner and CBT Guide performance in handling the responses from subjects during the three-stage CBT process (0: Valid, 1: Invalid). 139

5.1 Speci cations for the critical points on Volt-Var curve. 152

5.2 Speci cations for the critical points on Watt-Var curve. 154

5.3 Comparison of Estimation Accuracy between the Proposed Solver and power Solver for 30-Bus with 31 Unknowns. 162

5.4 Estimation Accuracy among 9-Bus, 30-Bus, 57-Bus, and 118-Bus Cases. 163

5.5 Number of buses with voltage within the prede ned range under three DRL methods and three grid service modes. 179

6.1 Summary of maximum vehicle range (R_{0G}), battery capacity (Q_0), estimated discharging rate (A_E), estimated capacity ratio (Q_E), and estimated charging rate (C_E) for the 10 most common electric vehicle types in NYC. 197

Acknowledgements

During my time at Columbia University, I have been incredibly fortunate to receive support from a remarkable group of individuals. First and foremost, I would like to extend my deepest gratitude to my advisors, Professor Xiaofan (Fred) Jiang and Professor Matthias Preindl. Their expansive vision, profound knowledge, insightful guidance, curiosity, and passion for research have been a constant source of inspiration. My journey with Fred began during my master's studies, and I consider myself extremely fortunate to have had him as one of my Ph.D. advisors. His friendship, unwavering support, and the countless lessons he has taught me—both within and beyond the realm of research—have profoundly shaped my growth as a researcher and scientist. I am equally grateful to Matthias for his patience, encouragement, and support as I navigated the intersections of power systems, machine learning, and human-centric optimization.

I would also like to thank Professor Nima Mesgarani, my rst guide into the world of computational neuroscience and neural network models for speech processing. My heartfelt appreciation also goes to Professor Xia Zhou, whose expertise in ubiquitous computing and inspiring leadership has been a source of motivation. I am also deeply grateful to Professor Gil Zussman for his guidance, encouragement, and mentorship. Outside of Columbia, I have had the privilege of collaborating with Professor Mohamed Abdel-Aty, whose interdisciplinary perspectives have provided invaluable insights, and Professor Xiaoyue Ni, who inspired me with how recent advancements in programmable matter and mechanical metamaterials could forward the capabilities of smart wearable technologies. I am also deeply grateful to Hanya Shao, a licensed psychotherapist, for her professional collaboration and her passion for exploring the intersections between embedded AI, sensing, and psychotherapy. Her insights and dedication have been invaluable to our interdisciplinary work.

I am deeply appreciative of my former and current colleagues who have stood by me throughout the ups and downs of research: Dr. Stephen Xia, Yanchen Liu, Dr. Liwei Zhou,

Minghui (Scott) Zhao, Yuan Fan, Kaiyuan Hou, Dr. Qijia Shao, Dr. Weizhong Wang, Dr. Matthew Jahnes, Youssef Fahmy, Dr. Peter Wei, Ziyi Xuan, Xinhua Sun, Yigong Hu, Runxi Wan, Dr. Haoxuan You and Bernard William Steyaert. I would also like to extend my gratitude to Professors Xiaofan Jiang, Matthias Preindl, Nima Mesgarani, and Debasis Mitra, and Dr. Vikram Mitra, who have generously contributed their time and expertise as members of my defense committee.

Along this journey, I have been incredibly fortunate to have the companionship of wonderful friends whose support has meant the world to me: Ke Wan, Xinyi Zhang, Ruoxi Liu, Xuechan Han, Ningrui Zhang, Xumei Mao, Ji Jia, Yian Yu, Dr. Danyang Zhuo, Dr. Boyuan Chen, Yinuo Qin, Dr. Jun Huang, Chun Hu, Dr. Cong Chen, Xinyu Cheng, Dr. Qi Yuan, and Jianquan Li. I would also like to extend my heartfelt thanks to my friends and fellow runners from the Newbee Running Club and Bergen Runners Running Club, whose enthusiasm and support have been instrumental to my running-related research.

I also had the privilege of spending two incredible summers as a machine learning research intern at Apple Inc., where I had the opportunity to work closely with Dr. Vikram Mitra, Dr. Erdrin Azemi, Dr. Behrooz Mahasseni, Dr. Ali Moin, Dr. Ran Liu, and Dr. Dung T. Tran. I am deeply grateful for the insights and knowledge I gained through our collaboration, which has had a lasting impact on my work.

I wish to express my deepest appreciation to my mother, Ming Xian, my father, Yijun Nie, my paternal grandfather, Guangqian Nie, my paternal grandmother, Zaoer Gao, my maternal grandfather, Guanghui Xian, and my maternal grandmother, Guizhen Chen—for their unwavering love, strength, and support throughout my life. I would like to express my special thanks to my spouse, Dr. Tingjun Chen, who continuously inspires and motivates me to become a better researcher. His unwavering support has been instrumental in helping me grow not only as a scholar but also as an independent woman.

Lastly, I am grateful to my beloved cats, Abby and Boss, who joined my life at the beginning of my Ph.D. journey and have lled my days with joy, comfort, and unconditional love,

and to our family dogs, Barbie and Chunchun, who kept my parents company in China and brought warmth and companionship to our family across the distance during these years.

Financial Support: The research presented in this thesis was supported in part by the Apple Scholars in AI/ML PhD Fellowship and by the National Science Foundation under Grant Numbers CNS-1704899, CNS-1815274, CNS-1943396, CNS-1837022, CMMI-2218809, and EEC-2133516.

Dedication

To my mom and dad, Ming Xian and Yijun Nie.

To my husband, Tingjun Chen.

Chapter 1: Introduction

The rise of connected devices and the Arti cial Intelligence of Things (AIoT), at the intersection of embedded arti cial intelligence (AI), wearable sensing, and intelligent edge computing, has opened new frontiers for advancing human wellness, physical health, and sustainable living. The evolution of sensing and computing technologies, coupled with advancements in embedded systems, is propelling forward the capabilities of wearable technologies, indoor sensing, and the development of smart and sustainable human-centric cities. These emerging capabilities are poised to reshape how we monitor health, deliver care, manage energy, and interact with the environments around us.

Yet, current solutions often fall short in accessibility, privacy, and sustainability. Many systems remain con ned to controlled environments, lack robustness in real-world deployment, or require costly infrastructures inaccessible to the broader population.

Motivated by the growing needs in mental health care, tness tracking, and sustainable energy use, this thesis investigates how embedded AI systems—especially when integrated into wearables, mobile platforms, and smart environments—can enhance human well-being at individual and community scales. This thesis embarks on designing systems and methods for hardware-software co-design at the intersection of embedded AI, wearable technology, and AIoT systems to enhance personal wellness, mental health, and environmental sustainability through sensing the physical world in its many forms. We aim to realize a future in which intelligent systems are seamlessly woven into daily life, proactively supporting individuals and communities in ways that are inclusive, private, and scalable.

By designing wearable devices and AIoT systems that blend naturally into people's routines, we tackle fundamental societal challenges—from emotion monitoring and running performance to mental health screening and sustainable EV charging. Using physics-informed signal process-

Figure 1.1: My research focuses on four interconnected thrusts spanning from the individual level to the system level: (Thrust 1) smart wearables and platforms for emotion, tness, and wellness [1, 2, 3, 4, 5, 6, 7, 8]; (Thrust 2) model-driven biosignal analysis and emotion detection [9, 10, 11, 12, 13]; (Thrust 3) AI-based mental and physical health care in smart home environments [14, 15, 16, 17, 18, 19, 20, 21]; and (Thrust 4) city-scale human-in-the-loop electric vehicle (EV)-interfaced grid optimization [22, 23, 24, 25, 26].

ing, multimodal data fusion, and edge-aware machine learning, we develop novel, resilient, and resource-ef cient solutions for future personal healthcare, indoor well-being, and city-scale energy optimization.

My research focuses on four interconnected thrusts spanning from the individual level to the system level, each addressing a different layer of human-centered intelligent sensing—from individuals to homes to cities as shown in Figure 1.1. This thesis presents contributions across four major topics across these four thrusts:

- Low-cost wearable systems for emotion recognition, focusing on real-time sensing of facial and ocular signals to detect apparent and real emotions in situ using infrared-based vision and physiological cues. (Thrust 1)
- Contactless acoustic sensing for tness and vital signs monitoring, introducing novel mobile-device-based methods to estimate key treadmill running metrics (cadence, ground contact time), and to analyze post-exercise physiological signals such as speech, breathing, and heart

sounds. (Thrust 1 and Thrust 2)

- AI-augmented smart home systems for daily functioning and mental health support, combining ambient sensing, psychotherapeutic conversation modules, and robotic agents to create a supportive home environment for individuals living alone or in need of care. (Thrust 3)
- Grid-aware EV-interfaced microgrid management and recommender systems, developing optimization and reinforcement learning frameworks that address the challenges of integrating electric vehicles and grid services in sustainable energy systems. (Thrust 4)

Together, these topic create a uni ed vision of embedded AI and sensing systems that are accessible, affordable, privacy-aware, and grounded in real-world utility. While each of the four thrusts focuses on distinct application domains, ranging from emotion recognition and tness tracking to mental health support and energy optimization, they are all guided by a common human-centric design philosophy. At the core of this thesis lies a uni ed, human-centric framework that begins with sensing and understanding the individual, whether through direct biosignal capture, interactive feedback, or behavioral data mining. By integrating embedded AI, machine learning, and multimodal sensing across personal, home, and urban environments, this research enables the analysis of users' habits, performance, and needs, and informs personalized recommendations and interventions to promote health, well-being, and sustainability in everyday life. In the remainder of this chapter, we provide an overview of each thrust and the major domains, as well as highlight their respective contributions to the literature.

1.1 Main Topics and Contributions

In this section, we present an overview of the main contributions of the main topics in this dissertation, guided by core principles of accessibility, pervasiveness, ef ciency, proactivity, affordability, and privacy awareness shown in Figure 1.2. These topics emphasize (i) system-level co-design across hardware, software, and AI models, (ii) privacy-aware and user-adaptive architecture design, and (iii) real-world deployments with diverse subject populations and environments.

Figure 1.2: The four topics of this thesis.

1.1.1 Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition

Wearable devices are driving a pivotal transformation in healthcare, shifting the approach from reactive treatment to proactive, continuous, and personalized care. In Chapter 2, to enable multiple applications where understanding human emotions is necessary, we present a System for Processing In-situ Bio-signal Data for Emotion Recog-nition and Sensing (SPIDERS+) – a low-cost, wireless, glasses-based platform for continuous in-situ monitoring of user's facial expressions (apparent emotions) and real emotions [1, 3]. We present algorithms to provide four core functions to detect eye shape and eyebrow movements, pupillometry, zygomaticus muscle movements, and head movements, using the bio-signals acquired from three non-contact sensors (IR camera, prox-

imity sensor, and IMU). SPIDERS+ distinguishes between different classes of apparent and real emotion states based on the aforementioned four bio-signals. We prototype advanced functionalities including facial expression detection and real emotion classi cation with a facial expression detector based on landmarks and optical ow that leverages changes in a user's eyebrows and eye shapes to achieve the average true positive rate up to 83.87%, as well as a pupillometry-based real emotion classi er with higher accuracy than existing low-cost wearable platforms that use sensors requiring skin contact. In addition, SPIDERS+ also includes optional contact-based sensors such as photoplethysmography (PPG) and multi-channel electroencephalogram (EEG) to monitor other bio-signals. SPIDERS+ costs less than $20 to assemble and can continuously run for up to 9 hours before recharging. We design, implement, and evaluate accompanying desktop and smartphone applications to demonstrate that SPIDERS+ is a truly wireless and portable platform that can impact a wide range of applications, where the knowledge of the user's emotional state is critical. We also showcased the potential of using this platform for applications in mental health monitoring and fatigue detection [2].

1.1.2 Contactless Acoustic-based System for Real-time Running Metrics Monitoring and Post-Exercise Body Signal Analysis

Running is one of the most widely accessible forms of physical activity, with approximately 50 million Americans participating in some form of running or jogging. It requires no expensive equipment, gym membership, coach, or team—making it an easy and independent way to stay active [27]. As a marathon enthusiast and certi ed running coach, I have engaged with countless runners through races, club events, and group runs. Many seek my advice on improving performance or avoiding injury, especially after learning about my background in electrical engineering, machine learning, IoT, and wearables. While professional lab tests using systems like Vicon provide accurate metrics, they are expensive and inaccessible to most runners [28]. Research alternatives—such as textile sensors or mmWave radar—often lack portability or practicality [29, 30]. Commercial wearables and smartphone apps are more accessible but include rigid compo-

nents that can cause skin irritation over long distances and often lack accuracy, particularly on treadmills. These challenges underscore the need for a more comfortable, accurate, and accessible solution for real-time treadmill running metric monitoring.

To address this need, in Chapter 4, we present SoundTrack, a contactless mobile solution that estimates these metrics by analyzing treadmill running sounds using on-device machine learning. Our main contributions are i) SoundTrackDB – a comprehensive 40-hour dataset of treadmill running sounds collected from 61 subjects across 363 sessions in 13 public gyms, created in collaboration with a licensed running coach; and ii) SoundTrack – an on-device mobile system capturing treadmill running sounds, mitigating noise, estimating cadence and GCT with a custom multi-layer perceptron (MLP) model, and providing real-time feedback. Microbenchmarks and evaluations show that SoundTrack effectively mitigates real-world noise challenges in public gyms and adapts to individual variations among runners and treadmill models. It achieves mean absolute percentage errors (MAPEs) of 1.62% for cadence and 6.05% for GCT on the test set of unseen running sessions, yielding results that are superior or comparable to commercial sports wearables. SoundTrack offers an accessible solution for treadmill metrics on mobile platforms, reducing reliance on specialized wearables and broadening accessibility [6, 5, 7].

In addition, cardio exercise, like running, elevates both heart rate and respiration rate, resulting in distinct physiological changes that affect speech patterns, pitch, breathing sounds, and heart sounds. These variations, which occur post-exercise, are in uenced by factors such as exercise intensity and individual tness levels. As such, in Chapter 4, in addition to SoundTrackDB, we also introduce a novel post-exercise body sound dataset consisting of over 250 sessions from 59 participants, capturing 143 minutes of structured reading, 47 minutes of spontaneous speech, 71 minutes of breathing sounds, and 62.5 minutes of phonocardiogram (PCG) data. Using this dataset, we conducted a case study on exertion-level classi cation based on post-run speech [13].

1.1.3 AI-based Mental and Physical Health Care in Smart Home Environments

Despite the growing global mental health crisis, the barrier to accessing mental health screening, professionals, and treatment remains prohibitively high. Leveraging AI and smart home advancements, in Chapter 4, in partnership with licensed therapists, we introduce CaiTI, a Care-home AI Therapist with Interventions, which integrates ambient sensing, natural language processing, robotics, and psychotherapeutic frameworks to support accessible mental health screening and intervention in smart home environments. Developed in collaboration with licensed therapists, CaiTI continuously monitors users' daily functioning across 37 dimensions using non-intrusive smart home sensors, and delivers personalized, evidence-based psychotherapeutic interventions using conversational AI. To simplify the use of smart home sensors for both users and professionals, we create a plug-and-play, mix-and-match sensor platform, featuring high con gurability and scalability [21]. It features two forms of intervention: natural language-based psychotherapy grounded in Motivational Interviewing (MI) and Cognitive Behavioral Therapy (CBT), and precautionary robot check-ins for physical wellbeing. CaiTI includes several novel technical contributions: (i) a reinforcement learning framework that adapts conversation ow based on user history and therapist-de ned priorities; (ii) integration of large language models (LLMs) with therapist-guided reasoning and validation mechanisms to ensure safety and personalization; and (iii) use of computer vision and vision-language models (VLMs) to detect concerning behaviors like smoking and drinking alone. The system was deployed in real homes with 20 users over 24 weeks, receiving positive feedback from both users and psychotherapists. To support reproducibility and future work, we open-source therapist-labeled datasets and task-speci c LLM prompts [14, 15, 17, 18].

1.1.4 Human-Centric Data-Driven Optimization and Recommendation in EV-Interfaced Grid at City Scale

Scaling up at the city scale, I am advancing human-in-the-loop electric vehicle (EV)-interfaced microgrids to promote sustainability. First, in Chapter 5, we focus on understanding the EV-

interfaced microgrids with standardized grid services. We address the challenges of estimating and optimizing grid behavior in EV-interfaced microgrids under partial observability. We propose a constrained optimization solver that accurately reconstructs grid states (active power, reactive power, voltage, and phase angle) from incomplete measurements, signi cantly outperforming existing methods such as pandapower and MATPOWER. We extend this solver to incorporate IEEE 1547-speci ed grid services (Volt-Var and Watt-Var modes), enabling the evaluation of their effects on grid stability and performance. To validate the solver, we conduct exploratory case studies across multiple IEEE bus systems (9, 30, 57, and 118-bus), demonstrating how different buses respond to service constraints and identifying critical nodes for optimization. In addition, we propose a deep reinforcement learning (DRL) approach using Twin Delayed Deep Deterministic Policy Gradients (TD3) to generate reactive control policies for DERs, which effectively regulate voltage uctuations in an IEEE 30-bus microgrid. This work contributes a novel tool for optimal power ow (OPF) estimation and control in DER-rich environments, helping utilities and stakeholders better integrate bidirectional EV charging, manage dynamic loads, and ne-tune grid service strategies.

Building on the top of the EV-interfaced microgrid optimizations and recognizing the impact of human behavior on the grid, in Chapter 6, we present a practical, data-driven, and human-centric EV charging recommendation system at the city scale, designed to optimize EV-grid interaction using DRL. This system DRL integrates user preferences, EV mobility, pricing, charger availability, and grid capacity, thereby co-optimizing welfare for EV drivers, charger owners, and grid operators. To address the challenges of large-scale EV integration, limited datasets, and complex spatiotemporal dynamics, we construct a rich EV-driving-charging dataset with ne-grained temporal resolution by aggregating public and self-collected data in Manhattan. Unlike most prior works that treat EVs as agents, our system treats EV chargers as agents, enabling more grid-aware and ef cient coordination of EV charging under real-world constraints such as grid capacity, pricing, and demand variation. Leveraging a deep Q-network (DQN) model, our framework learns to minimize EV-charger distances while optimizing load distribution and utilization [25, 26].

1.2 Contributions to Literature

The systems and approaches presented in this thesis have been evaluated through extensive real-world deployment, collaborations with clinical and engineering experts, and large-scale user studies. The work spans multiple disciplines—embedded systems, human-centered AI, human sensing, ambient intelligence, intelligent pervasive healthcare, and smart cities—and has resulted in publications across premier venues. Multiple contributions have received best paper and demo awards, and open-sourced datasets and tools developed in this thesis have supported further research across academia and industry.

In particular, the research on SPIDERS+ presented in Chapter 2 was published in the proceedings of ACM/IEEE IoTDI'20 with the demonstration of this platform received Best Demo Award at ACM/IEEE IPSN'20 [1, 3]. An extended version was published in Elsevier-Pervasive and Mobile Computing Journal (PMC) [2]. Additionally, we have open-sourced our glasses designs, code, and assembly instructions to the community to allow anyone to build upon or use SPIDERS+ for their own projects[1].

The preliminary results of the contactless acoustic-based system for real-time running metrics monitoring using the mobile device, presented in Chapter 3, were first presented as a workshop paper in ACM IASA'23 and a poster in ACM MobiCom'24 [5, 6]. The full system, SoundTrack, and the dataset curated, SoundTrackDB, were published in the proceedings of the ACM IMWUT'25 [7]. The work on the post-exercise body signal analysis was published in the proceedings of the ACM SenSys'25 [13]. We publicly release SoundTrackDB, the SoundTrack mobile app[2], and post-run body sound data[3] to democratize access to professional-grade running metrics and support broader research in acoustic sensing, fitness monitoring, and health-aware speech analysis.

CaiTI, the "AI therapist" work presenter in Chapter 4 was presented as a workshop paper under ACM IASA'22 [14]. This work offers an accessible, proactive approach to mental health care in everyday life, and a demonstration of the system received Best Demo Award Runner-up at

[1] https://github.com/Columbia-ICSL/SPIDERS
[2] https://github.com/Columbia-ICSL/SoundTrackDB
[3] https://github.com/Columbia-ICSL/data_after_cardio

ACM SenSys'22 [15]. The components of exploring foundation models in detecting concerning daily functioning in psychotherapeutic context based on images from smart home robots were presented as a workshop paper under IEEE FMSys'24 [18]. The LM-based conversational modules of CaiTI for daily functioning screening and psychotherapeutic interventions were published in ACM Health'25 [17]. The datasets prepared by the therapists and the few-shot prompts we designed with the therapists are publicly available.[4]

The research on optimal power flow (OPF) estimation for EV-interfaced microgrids with standardized grid services (Chapter 5), which mathematically models the dynamics of distributed energy resources (DER) and EVs, was published in the proceedings of IEEE ITEC'21, where it earned the Best Paper Award [22]. The extension version of this work was published in IEEE Transactions on Industry Applications'23 [24]. Another work in Chapter 5, the DRL-based approach for OPF with grid services, was published in the proceedings of IEEE/AIAA ITEC+EATS'22 [23].

The human-centric data-driven optimization and recommendation in EV interfaced grid at city scale work (Chapter 6) was accepted as a poster in ACM BuildSys'22 [25], and was subsequently published in the proceedings of ACM e-Energy'23 [26]. The dataset and framework in this work were released to support future research in large-scale EV mobility, charging infrastructure optimization, and sustainable urban planning.[5]

[4] https://github.com/Columbia-ICSL/CaiTI_dataset
[5] https://github.com/Columbia-ICSL/Data-Driven-Human-Centric-EV-Charging

Chapter 2: Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition

Intelligent emotion monitoring systems play an important role in improving the mental health conditions of the general public. They help inform the general population of their own mental states. They also provide psychiatrists and psychotherapists a means to monitor and diagnose their patients' mental health problems remotely. This is particularly critical during the COVID-19 global pandemic, as it is dif cult to schedule in-person and mask-off doctor visits. Emotion monitoring platforms also have immense impacts on businesses, providing companies with another measure of the effectiveness of their products and advertisements. Additionally, a low-cost and portable solution would enhance the lives of the general population by bringing, for example, enhanced entertainment experiences when movies/games/tv shows present dynamic content to users based on emotions experienced by the user in real-time and affordable mental health services to people who have access and cost barriers to mental health care [31, 32, 33, 34].

However, predicting emotions is challenging, let alone detecting emotions on a resource-constrained and wearable device. There are commercial products and research platforms developed for monitoring speci c signals that provide insight into a user's real emotion. However, many of the existing commercial products and sensors, such as EEG caps, are bulky, expensive, wired, and require constant skin contact, limiting their use to clinical or controlled environments. All of the existing vision-based approaches and products for detecting facial expressions utilize static features found within a single image to detect whether a person is smiling, sad, disgusted, etc. However, each person's eye and facial features, such as skin tone and eye shape, vary widely. This makes it dif cult to create an algorithm or train a classi er that can generalize to unseen examples well.

Figure 2.1: Left: The SPIDERS+ glasses platform including hardware sensors (red), PCB boards (blue), frame components (yellow), and optional attachments (green). Right: SPIDERS+ glasses on a mannequin.

Additionally, it is even more challenging to implement semi-real-time vision-based facial expression detectors on a resource-constrained, wearable, embedded platform. We observe that the facial and eye movements made by a person as he/she transitions into a new expression is more consistent across different demographics, and we leverage this information to develop our facial expression (apparent emotion) algorithm.

Pupillometry has been shown to be a strong indicator of real emotions [35, 36, 37]. However, even the state-of-the-art algorithms and systems that utilize pupillometry, such as the Tobii Pro Glasses 2, cannot reliably predict real emotions due to the inherent measurement noise caused by lighting changes, re ections, rapid eye movements, blinking, etc. As a result, in addition to real emotions, we provide facial expression (apparent emotion) as a direct output to support applications where apparent emotions are suf cient or preferred.

Our contributions in this chapter are summarized below:

- We present SPIDERS+, a low-cost, energy-ef cient, and portable wearable platform (see Figure 2.1) that collects data from a variety of non-contact sensors to detect: 1) eye & eyebrow shape, 2) pupillometry, 3) zygomaticus muscle (smile muscle) movements and 4) head movements. SPIDERS+ can also be extended to include functionalities provided by contact-

based sensors, such as 5) heart rate and 6) multi-channel EEG monitoring. We demonstrate the robustness of our sensing modalities by comparing them against commercial products.

- We present novel and robust vision-based techniques to perform pupillometry and eye & eyebrow shape detection using infrared (IR) band gray-scale images, from an IR camera positioned at a low angle from the eye as not to block the user's eld of view.

- We develop and implement a data processing pipeline that intelligently partitions tasks and computation across the glasses wearable and more powerful processing units (e.g., computer, cloud, smartphone) to maintain high accuracy, long battery life, and to ensure that our emotion recognition algorithms run in semi-real-time. We demonstrate that our wearable is able to last up to 9 hours.

- We propose a novel optical- ow-based algorithm for estimating facial expressions (apparent emotions) leveraging movements in the user's eyes as he/she transitions into a new facial expression. We experimentally show that our algorithm achieves an average classi cation accuracy of 83.87%, which outperforms the state-of-the-art approach (e.g., [34]). We also demonstrate SPIDERS+'s capability of estimating a user's affective state (real emotions), achieving a classi cation accuracy of 49.32%, which outperforms the results presented in [38].

- We demonstrate that SPIDERS+ can estimate user facial expressions and stream the results in semi-real-time to user-facing desktop and smartphone applications via WiFi [3]. This end-to-end pipeline can enable a wide range of user applications, such as fatigue detection and mental health monitoring.

- We provide the community with 1) an anonymized dataset of biometrics measured from subjects during our controlled experiments; 2) the circuit design and 3D model design for our eyeglasses wearable; and 3) a suite of algorithms and libraries that enable everyone to easily build their own emotion-based applications based on SPIDERS+.

2.1 Related Work

In this section, we first analyze related commercial products that track physiological data in order to identify issues in price, portability, and usefulness. We then analyze different modalities of data that would be useful for emotion classification purposes and explore algorithms that derive the types of data that cannot be directly measured by minimal sensors.

2.1.1 Bio-signal Acquisition by Wearable Devices

There are a variety of commercial bio-signal acquisition devices on the market, but all have specific problems that make them unsuitable candidates for a low-cost, energy-efficient, and portable platform for monitoring user emotions. For instance, Fitbit is a wireless wearable activity tracker, but only measures three-axis acceleration data, and photoplethysmography (PPG), both of which are not reliable measures of emotional states by themselves [39]. Muse is a headband that provides five-channel electroencephalogram (EEG) data, heart rate, and gyroscope sensors [40]. Both heart rate and EEG sensors require skin contact, which is difficult to maintain in a mobile wearable where the user may be on the move. Pupillometry is well-known as an indicator of affective states, and there are eye-tracking cameras on the market that measure these signals. For example, the Tobii Pro Glasses 2 leverages a wide-angle camera, microphone, gyroscope, and accelerometer to track the wearer's eye [41]. However, Tobii Pro Glasses 2 system is not truly wireless (requires an external recording unit) and is very expensive ($10,000+). To summarize, commercial products measure a limited amount of bio-signals that often require direct skin contact, or are very expensive and not portable.

A number of research platforms and glasses have been developed to provide a wearable platform for emotion monitoring. However, most of these systems focus exclusively on pupillometry [42, 43, 44], but not apparent emotion. The second category of works leverages photo-reflective IR sensors integrated within the eyeglass frames to detect facial expressions [45, 46]. However, using IR sensors provides limited resolution about the position of the cheeks/brows and no pupil-

lometry measurements, making it dif cult to perform robust facial expression recognition or real emotion classi cation across a varied population. [38] leverages sensor fusion, combining facial features extracted from a camera pointed at the user's eye and other physiological signals, such as skin conductance and PPG, to obtain a more accurate estimation of real emotions. However, we recognize that facial expressions are not reliable indicators of real emotions; for example, a user can easily fake a smile even if s/he is depressed. Additionally, measuring skin conductance and PPG requires direct contact between the skin and the sensor, which is not stable if the wearer is on the move.

2.1.2 Vision-Based Apparent Emotion Classi cation

Research about facial expression classi cation based on images of the whole face is abundant. Nevertheless, these approaches require continuous capturing of the user's entire face with a camera looking from a distance, which is not portable and can lead to privacy concerns. On the other hand, wearable devices are becoming more and more popular, and provide in-situ sensing capabilities [47, 48, 49, 50]. A camera mounted on the glasses can not only sense the pupil size and gaze angle, but can also clearly see the eyelids and eyebrows. Previous work directly uses raw images as input to a machine learning classi er [38]. However, directly learning the features with raw images requires a large amount of labeled data, and it can perform well within the dataset but poorly in real-world cases.

For eyelid and eyebrow detection, many previous works utilize RGB images, conducting adaptive thresholding, luminance valley detection, and edge detection to detect the shape of the eyelids and eyebrows [51, 52, 53]. However, these methods are computationally expensive and not robust. Also, in order to perform pupillometry, we have to use IR cameras that produce gray-scale images, making the boundary between the sclera (white outer layer of the eye) and skin harder to distinguish. Recent works also focus on leveraging IR cameras to conduct deep-learning-network-based facial expression classi cation on resource-constrained platforms. However, the hardware form factors of many of these systems are bulky and not comfortable for daily wear [54].

Convolutional neural networks (CNN) and regression tree models have been commonly used to generate facial landmarks for facial expression recognition. These methods estimate facial landmark points that draw out the location and curvature of the eyelids and brows [55, 56], and some of these methods have been shown to run in real-time [57]. As such, we decide to adopt a similar approach in landmark generation. However, generating the shapes of the eyelids and brows and determining the user's expression and apparent emotions based on this information are two different problems, the latter of which has not been addressed in previous works.

2.1.3 Real Emotion Recognition

A variety of bio-signals have been explored to recognize a person's real emotional state. Pupillometry is one of the most popular signals used to classify real emotional states due to its non-invasive nature and its high discriminatory power between different classes of real emotion. To achieve high accuracy in classifying arousal and valence levels (two measures of real emotions), the majority of works in the literature acquire the pupillometry and gaze information from commercial eye trackers that are expensive and not truly wireless [58, 36, 59].

Other bio-signals that have been linked to real emotions are heart rate and EEG. Long-term heart rate variability (HRV) has been shown to be correlated to emotional patterns and shifts [60, 61]. EEG head caps are also widely used in detecting emotion because they non-intrusively capture brain signals [62, 63, 64]. However, both heart rate monitoring and EEG monitoring require sensors to be in direct contact with the skin; in a wearable system, where a user could be moving, maintaining contact is not guaranteed nor very comfortable for the wearer. Additionally, EEG-based methods have achieved high-performance rates because of the sheer number of channels employed in these works (often 32 or more). It is difficult to incorporate even a few channels of EEG into a resource-constrained wearable.

2.2 SPIDERS+System Design

SPIDERS+ is a low-cost, long-lasting, and mobile platform for monitoring apparent and real

Figure 2.2: The four-layer system architecture of SPIDERS+. We designed the architecture in a top-down approach based on the types of applications that SPIDERS+ aims to impact. All components and layers in SPIDERS+ work towards enabling or improving the top layer Enabled Applications.

emotions of the wearer. In this section, we present our hardware design and data processing architecture that allows SPIDERS+ to achieve robust apparent and real emotional recognition while being energy-efficient and maintaining a suitable latency.

Figure 2.2 shows the functional system architecture of SPIDERS+. SPIDERS+ focuses on prototyping and enabling its functionalities and applications based on inputs from non-contact sensors, while also providing extended features and functionalities by enabling more intrusive, but optional, contact-based sensors. Each layer in the architecture directly enables the applications or functionalities in the above layers. We take a top-down approach in designing SPIDERS+ based on the types of applications we aim to address. In particular, the Hardware Sensors layer, consists of a wide range of contact-less and optional contact-based sensors that SPIDERS+ supports. SPIDERS+ uses outputs from its Hardware Sensors to extract a common set of bio-signal measurements, such as pupillometry and eye shape, that form the Core Function Library. The measurements that are extracted in the Core Function Library are all bio-signals that are correlated with a user's emotional state. These features are then utilized by the Advanced Functionalities layer, that runs novel machine learning and signal processing algorithms (Section 2.4), to estimate aspects of user emotion, including facial expressions and

internal emotion. Outputs from the Advanced Functionalities layer are also the outputs of SPIDERS+, but can be used to improve and enable a multitude of applications as illustrated in Section 2.5.

There are many applications and problems that untethered emotion monitoring could enable; our aim is to provide the community with a platform that enables emotion monitoring and allows researchers to develop their own algorithms to address a wide range of applications by leveraging the plethora of sensors and modalities that SPIDERS+ provides. In addition, we recognize that there are two facets of emotions that people exhibit, apparent and real emotion, provided at the Advanced Functionalities layer. By providing a mobile platform that is capable of measuring these two dimensions of emotion, we believe that we can impact some of the Enabled Applications shown in Figure 2.2. We also develop user-facing desktop and smartphone applications that leverage outputs from SPIDERS+ Apparent Facial Expression Detection module to enable applications such as real-time apparent emotion monitoring and fatigue detection.

2.2.1 Sensing Modalities Overview

We now discuss the types of bio-signals incorporated in SPIDERS+ based on what is needed to achieve robust apparent and real emotion recognition. Pupillometry, EEG, HRV, skin conductance, and facial expressions are commonly used to determine a person's emotional state, as discussed in Section 2.1. Previous works generally leverage all or a subset of these signals to classify emotional states. First, we recognize that there are two facets of emotion that people exhibit: apparent and real emotion, and that knowing one does not necessarily provide the other. For example, people may hide their internal feelings of sadness by smiling if he/she is celebrating or spending time with friends and family.

Common signals used to determine someone's real emotions include pupillometry, skin conductance, EEG, and HRV. We omit facial expressions, as people's expressions may not accurately represent their real emotions. Skin conductance, EEG, and HRV are all commonly measured using sensors that require skin contact. However, in a mobile wearable system, the contact between the

sensor and skin can often be unstable, making these signals difficult to acquire. Pupillometry has been shown to be powerful in discriminating between different emotional states and can be measured using a camera, which is non-contact. The IR-filtered camera coupled with an IR illuminator usually performs better in conducting pupillometry than the commonly used RGB cameras. Hence, we integrate an IR-band gray-scale camera into SPIDERS+ to perform pupillometry to classify real emotion. Our proposed methods for obtaining robust pupillometry in the face of these challenges are described in Section 2.3.2.

Facial expressions are representative of a person's apparent emotion, or the emotion that people express through visible body language. Facial expressions, head movements, and zygomaticus muscle movements can be adequately captured using a camera, inertial measurement units (IMU), and proximity sensors pointed at a person's face. Both are non-contact sensors, and proximity sensors have negligible power consumption when compared to that of a camera. Our novel algorithms for determining facial expressions and apparent emotion using signals from these sensors are detailed in Section 2.3.1 and Section 2.4.

A summary of the different bio-signals that SPIDERS+ provides is shown in the Core Function Library layer in Figure 2.2. A summary of the different sensors integrated into SPIDERS+ to enable the measurement of these bio-signals can be found in the Hardware Sensors layer.

2.2.2 Hardware Platform

In this section, we describe the hardware platform implementation of SPIDERS+ that allows us to perform robust apparent emotion detection and real emotion classification efficiently and with low latency. Additionally, we have open-sourced our glasses designs, code, and assembly instructions to the community to allow anyone to build upon or use SPIDERS+ for their own projects.[1]

[1] https://github.com/Columbia-ICSL/SPIDERS

Processing and Wireless Communication Components

The system on a chip (SoC) module is responsible for sampling and transmitting sensor data from the wearable glasses to the client application. A microcontroller with proper performance is required to sample and transmit data from all the sensors with enough temporal resolution. To minimize the PCB layout area, we also prefer wireless SoC modules with an integrated radio module. Based on these requirements, we chose the Espressif ESP32 as our processor and wireless transmission module, which supports both WiFi and Bluetooth communication protocols, computational power at low power consumption, and direct memory access (DMA).

Hardware Sensors

The camera sensor captures eye & eyebrow shape as well as pupillometry. We chose the OV2640 camera module because of its high sampling rate and moderate pixel resolution (320 240). Additionally, the OV2640 module performs on-chip JPEG encoding, converting pixels into a machine-readable format without computation required from the MCU. We also install an 850 nm band-pass IR lter and two IR LEDs in front of the camera lens to illuminate the areas directly in front of the camera and to eliminate the in uence of different iris colors and ambient light from our images.

Compared to the camera, the proximity sensor and IMU are relatively compact and low-power. We decided to use a VL6180X proximity sensor to measure the distance between the sensor and the zygomaticus muscle (smile muscle) and an MPU6050 gyroscope and accelerator sensor to detect the head movements.

In addition to the non-contact sensors that are used to extract features to estimate facial expressions and internal emotions in the Advanced Functionalities layer, SPIDERS+'s hardware platform also supports several optional contact-based sensors to measure heart rate and EEG. In particular, SPIDERS+ uses the MAX30102 photoplethysmography (PPG) sensor for heart rate detection and also incorporates the TI ADS1298, a low-power, high-accuracy 8-channel ADC for sampling EEG signals. SPIDERS+ is capable of maintaining a sampling rate of 250 Hz across all