MULTI-SCALE CONTACTLESS SENSING: ADAPTIVE SIGNAL PROCESSING AND FEATURE FUSION FOR ROBUST RECOGNITION

JUNCEN ZHU

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University
Department of Computing

Multi-Scale Contactless Sensing: Adaptive Signal Processing and Feature Fusion for Robust Recognition

Juncen Zhu

A thesis submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
April 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

Signature: ____________________

Name of Student: Juncen Zhu

Abstract

With the continuous advancement of artificial intelligence (AI) and the Internet of Things (IoT), contactless sensing has emerged as a pivotal technology, offering non-intrusive, high-resolution monitoring across diverse applications. However, ensuring robust real-time sensing in dynamic and complex environments remains a significant challenge. Current systems face three major bottlenecks. First, at the signal preprocessing stage, multipath interference and noise degrade the fidelity of weak signal components, undermining the quality of downstream analysis. Second, during feature extraction, signals originating from multiple spatial regions and hierarchical levels become intertwined with irrelevant information, making it challenging to isolate salient features critical for accurate recognition. Finally, in the feature fusion stage, conventional single-scale approaches struggle to integrate transient fluctuations with long-term trends, limiting the system's ability to perform effective multi-modal and cross-temporal data fusion.

To address these challenges, this study develops a contactless sensing framework that integrates adaptive signal processing and feature fusion to enhance robustness, precision, and scalability. First, we propose an adaptive signal processing pipeline, incorporating a parametrically tunable windowing function and frequency-domain transformation strategy. This method dynamically balances spectral resolution and side-lobe suppression, ensuring high-fidelity signal acquisition even in low signal-to-noise ratio (SNR) conditions.

For feature extraction, we introduce a hierarchical feature selection and refinement mechanism, which iteratively enhances target signals while mitigating background interference. Furthermore, a cross-domain feature fusion strategy is employed, leveraging generative projection and domain-adaptive reconstruction to reduce noise sensitivity while preserving discriminative features. Compared to conventional filtering or threshold-based methods, this approach eliminates anomalies dynamically without relying on static parameters or pre-defined priors, ensuring generalizability across diverse sensing conditions.

To simultaneously capture short-term variations and long-term trends, we propose a multi-resolution temporal fusion strategy, aligning information across different time scales into a unified analytical model. This mitigates the limitations of single-scale processing, allowing the system to maintain robust recognition performance in dynamically evolving environments, particularly for physiological state monitoring.

By integrating these methodologies, the proposed sensing system achieves real-time efficiency, high-resolution signal processing, and robust recognition under complex, multi-source interference conditions. Ultimately, this study establishes a scalable, adaptive, and high-precision contactless sensing paradigm, advancing its practical applicability in real-world intelligent sensing and monitoring systems.

Publications Arising from the Thesis

1. Juncen Zhu, Jiannong Cao, Yanni Yang, Wei Ren, and Huizi Han, "mmDrive: Fine-grained Fatigue Driving Detection Using mmWave Radar", in *ACM Transactions on Internet of Things*, Vol. 4, No. 4, Article 26, November 2023.
2. Juncen Zhu, Huizi Han, Jiannong Cao, Julie McCann, Ho-Yin Michael Ma, and Xiaoyun Liu, "LiquidAuth: Reliable and Accurate Liquid Authentication Using GAN-enhanced Acoustic-to-Mass-Spectrum Mapping", in *Proceedings of the 34th International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–9, IEEE, 2025. (accepted)
3. Juncen Zhu and Jiannong Cao, "Robust mmWave Radar Sensing via PSLR-Adaptive Windowing and Noise-Aware GED-MUSIC", manuscript ready.

Acknowledgments

As I reflect on this transformative journey spanning four years of doctoral studies, I am filled with profound gratitude for the remarkable growth I have experienced. This journey has not only deepened my academic knowledge but has fundamentally shaped me into a more mature and resolute individual, better equipped to face life's challenges with confidence and wisdom. The person I am today—more thoughtful, more determined, and more capable of navigating complexity—is a testament to the invaluable guidance and support I have received along the way.

I would like to express my deepest gratitude to my supervisor, Professor Jiannong Cao, whose mentorship has been instrumental throughout my entire doctoral experience. His guidance has been invaluable not only in academic learning and research work, but also in cultivating my capabilities as a researcher and developing my comprehensive skills. Time and again, when I encountered difficulties or felt lost, Professor Cao provided unwavering affirmation, recognition, and substantial practical support that guided me forward. I am also deeply grateful to Professor Julie McCann from Imperial College London, who served as my supervisor during my visiting period and provided invaluable insights that enriched my research perspective.

My heartfelt appreciation goes to Dr. Yanni Yang, whose mentorship has been pivotal in my journey in the field of wireless sensing. She taught me everything from the initial experimental coding to getting started in this research area. Her incredibly important guidance not only shaped my early research foundation but continues to influence my

learning and development. I also extend my sincere thanks to Dr. Wei Ren, Dr. Mingjin Zhang, Dr. Tao Wu, and Dr. Linfei Ge, who provided tremendous assistance in helping me organize my thoughts and navigate various challenges throughout my studies. Their wisdom and support have been invaluable to my academic growth.

I am grateful to all my colleagues in the IMCL, whose weekly meetings and discussions not only provided me with direction for moving forward but also created a warm, family-like atmosphere that made this journey meaningful. Special thanks to Dr. Yu Yang, my laboratory colleague, and to my wonderful classmates Rui Cao, Yuqing Zhao, and Yingfeng Cao, who have been constant sources of support and friendship.

Finally, I owe my deepest gratitude to my family and closest friends who sustained me through the most vulnerable moments of this doctoral journey. To my parents, Mr. Yongqi Zhu and Ms. Kailu Tang, whose unwavering love and support have been my anchor. To my IMCL colleague and roommate Ms. Xiaoyun Liu, and my dear friends Ms. Yu Gao and Ms. Can Gao, who provided steadfast encouragement, support, and care during times when I felt confused about my research and life direction. Their presence and belief in me made all the difference in completing this journey.

This achievement belongs not only to me but to all of you who walked this path alongside me. Thank you for making this dream possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

In fields such as healthcare, intelligent transportation, security monitoring, and industrial inspection, sensing technology plays a crucial role in acquiring environmental information and monitoring target status. Traditional contact sensing technologies, such as electrode-based physiological monitoring, wearable biosensors, and pressure sensors, are widely used due to their high stability and precision. These methods have been extensively applied in vital sign monitoring, driver status assessment, and intelligent manufacturing. However, their dependence on direct physical contact introduces inherent limitations in terms of long-term monitoring, user comfort, privacy protection, and environmental adaptability. For example, long-term use of wearable devices may cause discomfort, camera-based monitoring raises privacy concerns, and physical contact sensors are susceptible to external environmental factors such as humidity and skin conditions. Consequently, non-contact sensing, which eliminates the need for direct contact while maintaining effective perception, has received increasing attention and demonstrates significant potential in remote health monitoring, intelligent interaction, behavior analysis, and industrial quality control.

Non-contact sensing utilizes the reflection, scattering, and attenuation characteristics of wireless signals to achieve remote status monitoring without requiring the monitored object to wear any device. Compared to traditional contact methods, this approach not only reduces user burden and improves the convenience of long-term monitoring but also provides unique advantages in privacy protection, long-distance detection, and environmental adaptability. In recent years, advances in related technologies such as radio frequency (RF) sensing and acoustic sensing have expanded the application scope of non-contact sensing. For instance, in the field of intelligent driving, millimeter-wave radar detects driver fatigue by capturing precise head and eye movement patterns, providing reliable safety warnings. In high-resolution radar sensing, researchers have optimized adaptive window functions to enhance target detection in complex environments. Additionally, in the field of liquid authentication, acoustic-based detection methods combined with deep learning achieve non-destructive and efficient liquid identification, offering new technical solutions for food safety and industrial inspection.

Despite significant progress in non-contact sensing across multiple domains, existing systems still face major challenges. At the signal preprocessing stage, complex environments and dynamic target characteristics introduce multipath interference and background noise, making it difficult for existing methods to effectively suppress these interferences. This limitation constrains the system's ability to detect subtle changes. For example, traditional signal processing methods typically rely on fixed window filtering techniques which, although effective in noise reduction to a certain extent, lack adaptability to dynamically changing environments. Consequently, these methods struggle to maintain an optimal balance between noise suppression and feature preservation. Furthermore, at the feature extraction stage, signals typically originate from multiple spatial regions and hierarchical levels, accompanied by substantial irrelevant information. Distinguishing key information while suppressing background noise poses a considerable challenge to system robustness. Existing methods primarily rely on fixed threshold decision rules or predefined feature selection, lacking adaptability to different environmental conditions, thereby reducing system stability in practical applications.

Moreover, at the pattern recognition stage, different application scenarios have different requirements. Some tasks emphasize the detection of short-term transient signals, while others rely on long-term trend analysis. However, many non-contact sensing methods still employ fixed-scale feature analysis, making it difficult to capture both transient and persistent patterns simultaneously. This limitation hinders their adaptability, particularly in multimodal and cross-temporal data fusion. Therefore, optimizing the three fundamental components of signal preprocessing, feature extraction, and feature fusion is crucial for improving the overall performance of non-contact sensing systems.

1.2 Non-contact Wireless Sensing System Framework

Given these challenges, the performance of non-contact wireless sensing systems depends on the end-to-end processing pipeline from signal acquisition to pattern recognition. Each stage of this pipeline is crucial for ensuring accuracy, robustness, and adaptability in complex environments. Therefore, this research further analyzes the system framework of non-contact wireless sensing and explores key processing strategies at each stage.

An overview of the general non-contact wireless sensing system architecture is illustrated in Fig. 1.1, highlighting the five key processing stages and their associated techniques and challenges.

A standard non-contact wireless sensing system typically consists of five key stages:

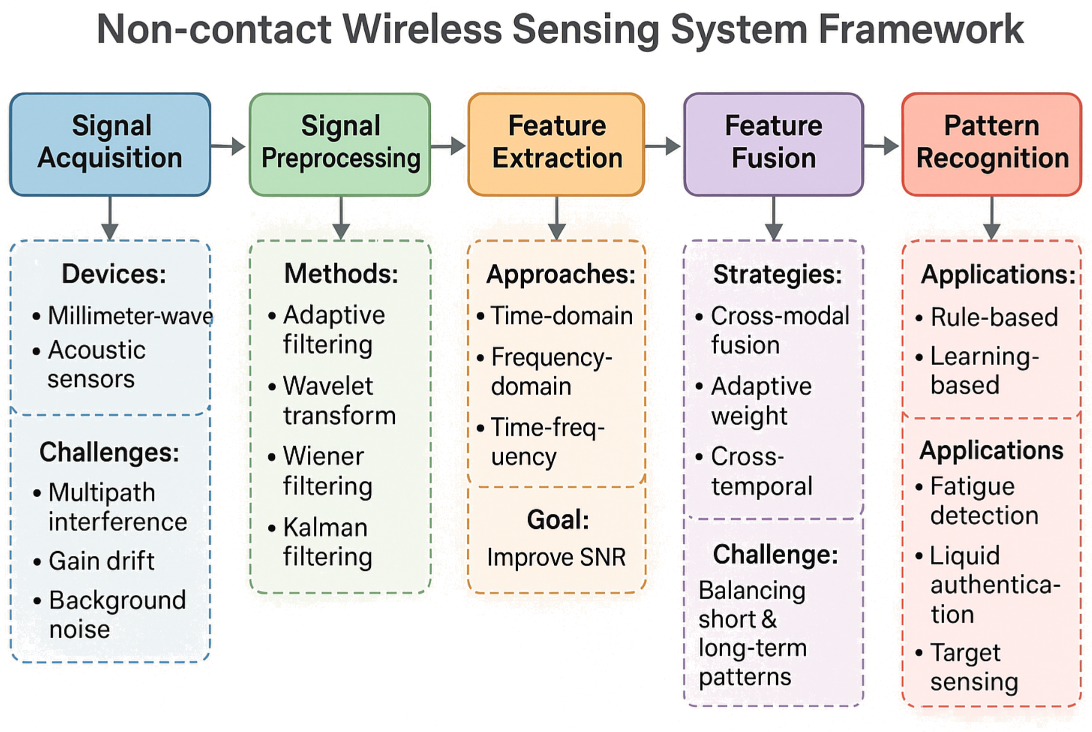


Figure 1.1: Overview of a non-contact wireless sensing system, illustrating the processing pipeline from signal acquisition to pattern recognition, along with key methods and challenges at each stage.

- **Signal Acquisition:** Wireless sensing devices, such as millimeter-wave radar and acoustic sensors, capture target signals from the environment. However, these signals are typically affected by multipath interference, gain drift, and background noise, complicating subsequent processing. To mitigate these issues, advanced hardware solutions, including high-gain antenna design and multi-antenna arrays, are employed. Additionally, Automatic Gain Control (AGC) and real-time calibration techniques enhance signal quality, providing stable input for further processing.

- **Signal Preprocessing:** Wireless signals are susceptible to environmental noise, dynamic interference, and multipath effects, necessitating robust preprocessing techniques. Traditional preprocessing methods include fixed-parameter filtering methods (such as low-pass, high-pass filters), while modern adaptive techniques such as wavelet transform, Wiener filtering, and Kalman filtering dynamically adjust denoising strategies to improve the signal-to-noise ratio (SNR).

- **Feature Extraction:** Extracting information-rich representations from preprocessed signals is critical for accurate state recognition. Feature extraction methods can be categorized into physics-based approaches (such as time-domain, frequency-domain, and time-frequency features) and data-driven approaches (such as principal component analysis, autoencoders). For example, in liquid authentication, spectral features of acoustic signals provide high discriminative power.
- **Feature Fusion:** In complex scenarios, a single feature may not be sufficient for comprehensive sensing. To enhance sensing accuracy, systems employ cross-modal fusion (such as combining mass spectrometry with acoustic sensing) or adaptive weight fusion (dynamically adjusting feature importance based on feature correlation). For instance, in driver fatigue detection, integrating short-term eye movement patterns with long-term head posture changes improves the accuracy of state estimation.
- **Pattern Recognition:** Extracted features are used for classification or regression using rule-based and learning-based approaches. Rule-based methods rely on threshold-based and logical rule evaluation, for example, in driver fatigue detection, when the blinking frequency exceeds a predefined threshold or the head tilt angle is too large and persists for a period of time, a fatigue alert is triggered. In contrast, learning-based methods leverage machine learning and deep learning to automatically learn decision boundaries and feature correlations.

While existing non-contact wireless sensing approaches have demonstrated promising results in various applications, several key challenges remain. At the signal preprocessing stage, suppressing multipath interference remains difficult, leading to degraded signal quality. At the feature extraction stage, distinguishing relevant target signals from background noise remains a major challenge. At the feature fusion stage, ensuring adaptability in dynamic environments and addressing individual variations remains an open issue. Furthermore, in cross-temporal analysis, most sensing approaches still rely on fixed time window methods, lacking the ability to simultaneously model short-term dynamics and long-term trends, thereby affecting generalization capability in complex scenarios.

1.3 Research Framework and Scope

Non-contact wireless sensing systems aim to sense target states and behaviors without requiring direct physical contact, achieving robust operation in complex dynamic environments. However, existing methods still face challenges in signal preprocessing, feature extraction, and feature fusion, limiting their adaptability and generalization capabilities. At the signal preprocessing stage, multipath interference, background noise, and dynamic characteristics of the target environment degrade signal quality. Traditional filtering and denoising methods struggle to suppress noise while preserving key information. At the feature extraction stage, signals typically originate from multiple spatial regions or hierarchical levels, accompanied by substantial irrelevant components. Traditional methods based on fixed thresholds or static feature selection lack adaptability to different environments, reducing system robustness. Additionally, at the feature fusion stage, existing methods primarily rely on single-scale temporal analysis, unable to effectively integrate information from different time scales, which limits their performance in multimodal and cross-temporal data fusion tasks.

An overview of the proposed research framework is shown in Fig. 1.2, highlighting the optimization designs made at different stages to improve robustness, accuracy, and generalization capabilities.

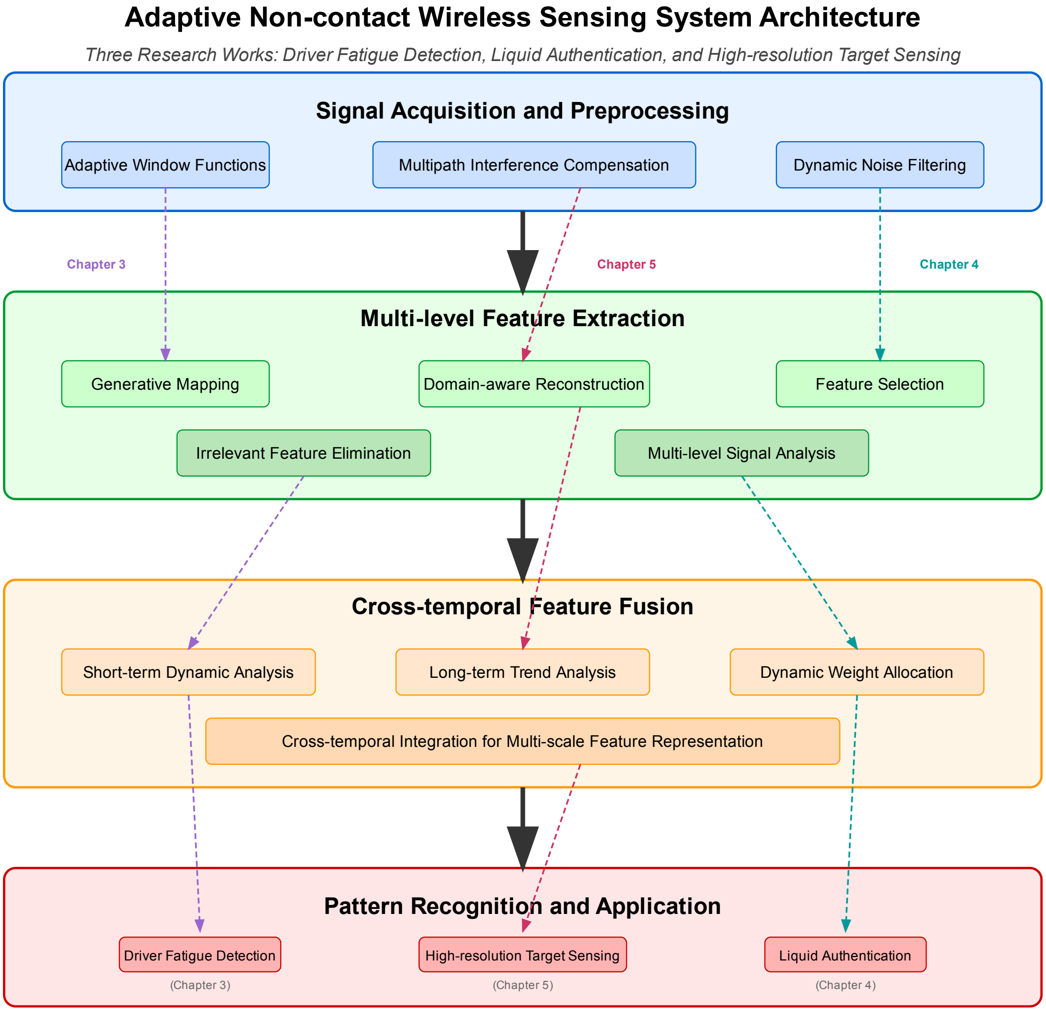


Figure 1.2: Adaptive non-contact wireless sensing system architecture, illustrating the key challenges and optimization strategies across signal preprocessing, feature extraction, feature fusion, and application stages.

To improve signal preprocessing, we propose adaptive window functions and frequency domain transformation strategies that dynamically adjust parameters to enhance signal quality under complex conditions. Unlike traditional fixed window techniques, this approach optimizes the balance between main lobe resolution and side lobe suppression, improving weak signal detection even in low signal-to-noise ratio (SNR) scenarios. Furthermore, we integrate multipath interference compensation mechanisms that adaptively adjust parameters in dynamic environments, ensuring reliable signal representation and enhancing the stability of subsequent feature extraction.

For feature extraction, we introduce multi-level feature analysis and dynamic selection mechanisms to address challenges caused by overlapping signals and background noise. The approach employs generative mapping and domain-aware reconstruction to project the original signal into a latent space insensitive to noise and interference, then emphasizes significant features through the reconstruction process. Compared to direct filtering or fixed threshold techniques, this strategy adaptively eliminates artifacts and distortions while minimizing dependence on predefined priors, thereby maintaining high robustness and accuracy across different environments. Additionally, to mitigate the interference of non-target movements, we integrate custom irrelevant feature elimination algorithms to separate unwanted motion signals, ensuring more reliable feature extraction.

For feature fusion, we propose a method capable of integrating information from different time scales, enabling the system to analyze signal characteristics more comprehensively. This approach, through dynamic weight allocation, balances contributions from various time scales, ensuring that both short-term changes and long-term trends are considered in the decision-making process, thereby improving system generalization capability in complex environments.

By integrating these optimization strategies, the proposed method achieves significant improvements in signal preprocessing, feature extraction, and feature fusion. Experimental results demonstrate that the proposed method enhances performance in various tasks, including driver fatigue monitoring, high-resolution target sensing, and liquid authentication. The research findings validate the robustness, accuracy, and adaptability of the proposed framework, highlighting its potential for broader applications in practical scenarios.

1.4 Thesis Organization

The remainder of this thesis is structured as follows:

Chapter 2 provides a comprehensive literature review on non-contact sensing technologies, including radio frequency (RF) sensing, acoustic sensing, and other non-invasive methods. This chapter highlights the strengths and limitations of current approaches and identifies key research gaps that drive this study.

Chapter 3 introduces a millimeter-wave radar-based fine-grained driver fatigue detection system. This chapter elaborates on the challenges of fatigue detection in driving scenarios and proposes a non-contact sensing framework capable of extracting physiological and behavioral fatigue indicators. System design, signal processing techniques, and multimodal fusion strategies are discussed, along with performance evaluation in real driving environments.

Chapter 4 presents a novel liquid authentication system that integrates acoustic sensing with deep learning spectral mapping. The system leverages Generative Adversarial Networks (GANs) to map acoustic absorption-transmission characteristics to mass spectrometry features, enabling precise and non-destructive liquid identification. This chapter details the system architecture, signal processing techniques, and experimental validation.

Chapter 5 introduces adaptive and robust window function design for high-resolution millimeter-wave radar sensing. This chapter addresses the fundamental trade-off between main lobe resolution and side lobe suppression and proposes dynamic optimization strategies to improve object detection in complex environments. The effectiveness of the proposed method is validated through simulations and actual experiments.

Chapter 6 discusses the significance and broader impact of our research on non-contact wireless sensing systems. It synthesizes advancements in signal preprocessing, feature extraction, and temporal fusion across our three applications. The chapter examines theoretical contributions, practical societal impacts in transportation safety and healthcare, and addresses current limitations in environmental adaptability, individual variability, and computational requirements.

Chapter 7 presents the dissertation's conclusions, summarizing core contributions in feature granularity and differentiation techniques across liquid authentication, driver fatigue detection, and anxiety monitoring applications. It highlights both theoretical advances and practical implications while identifying promising future research directions in scalability, edge computing integration, and interdisciplinary applications.

Chapter 2

Literature Review

With the rapid advancement of wireless communication technologies, contactless sensing has emerged as a crucial enabler across various domains, facilitating efficient and non-intrusive solutions for human activity monitoring, driver state assessment, and material identification. Today, a diverse range of wireless sensing devices is widely adopted, including millimeter-wave (mmWave) radar, ultra-wideband (UWB) radar, and ultrasonic sensors for environmental perception and target detection. Additionally, embedded acoustic sensing capabilities in consumer electronics, such as microphones and speakers in smartphones, further extend the reach of contactless sensing applications. The increasing ubiquity of wireless signals has paved the way for advancements in contactless sensing, establishing it as a foundational technology in fields such as healthcare, intelligent transportation, and food safety.

An overview of the typical methods, challenges, and emerging research trends in contactless sensing is illustrated in Fig. 2.1. This figure highlights the motivations driving the development of new frameworks across three core application areas.

Over the past decade, researchers have actively investigated the application of wireless signals in contactless sensing, with a focus on enhancing precision, robustness, and real-time adaptability. Radar-based sensing has been extensively utilized for phys-

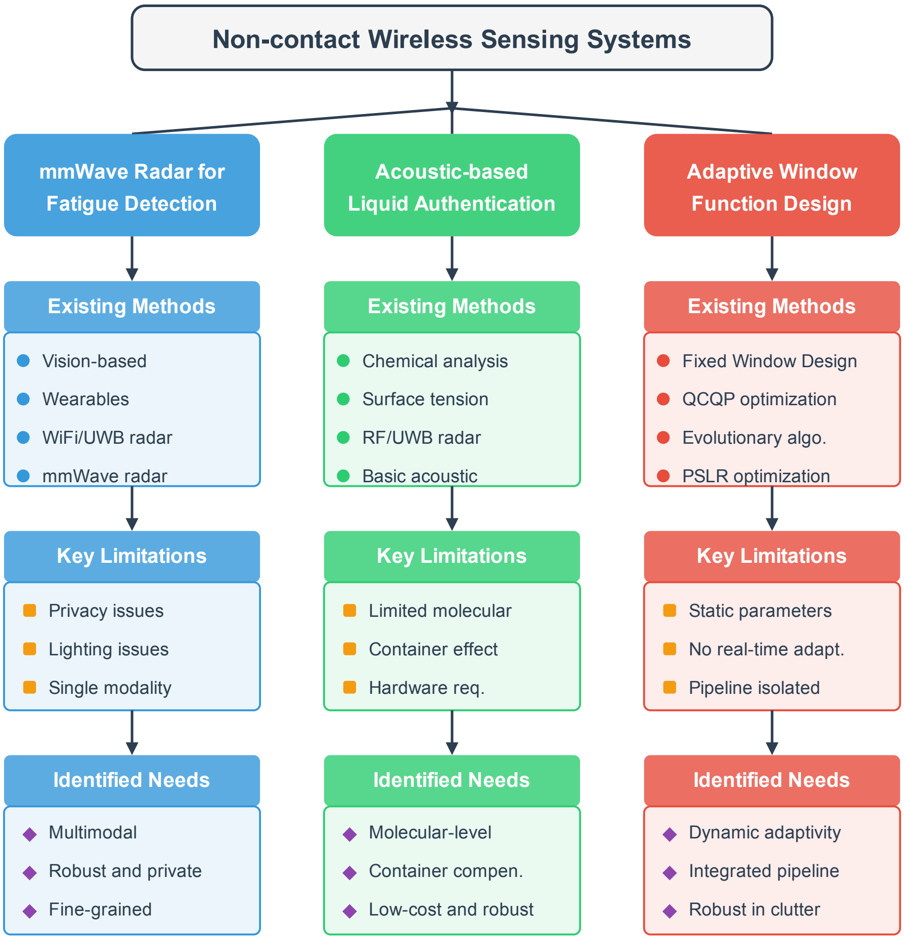


Figure 2.1: Summary of typical contactless sensing methods, challenges, and targeted improvements in three application areas: mmWave radar-based fatigue detection, acoustic-based liquid authentication, and adaptive window function design.

iological signal monitoring, fatigue detection, and behavioral recognition, whereas acoustic-based sensing has demonstrated efficacy in liquid authentication and material identification. Despite significant progress in controlled environments, practical deployments still face substantial challenges in dynamic and cluttered scenarios, where signal interference, background noise, target motion, and multi-scale information integration complicate sensing reliability.

This chapter provides a comprehensive review of existing contactless sensing methodologies and contextualizes them within the three core application areas addressed in this dissertation: mmWave radar-based fatigue driving detection, acoustic-based liquid authentication, and adaptive window function design for high-resolution mmWave radar sensing. We begin by outlining the fundamental principles and enabling technologies of contactless sensing. We then examine state-of-the-art techniques in each application area, analyzing their methodological advancements, inherent limitations, and persisting challenges. This review serves as the theoretical foundation for the proposed methodologies in subsequent chapters.

2.1 Research Progress in mmWave Radar-based Fatigue Driving Detection

Fatigue driving is a major cause of traffic accidents, making accurate and reliable detection methods essential for enhancing road safety. Over the years, researchers have explored various sensing modalities for fatigue detection, including vision-based, wearable-based, and RF-based approaches. While vision-based and wearable-based solutions have shown promise, they present inherent limitations in real-world driving scenarios. In contrast, RF-based approaches, particularly mmWave radar, have emerged as a robust alternative due to their high resolution, robustness to lighting conditions, and capability to capture both facial and physiological fatigue features.

This section reviews the advancements and limitations of these approaches and highlights the potential of mmWave radar for fatigue detection.

2.1.1 Vision-based Approaches

Vision-based approaches primarily rely on cameras and infrared (IR) sensors to capture facial expressions and eye movements indicative of driver fatigue. Features such as nodding, yawning, eye blinking, and gaze direction have been extensively studied using machine learning algorithms, including back-propagation neural networks (BPNN) and convolutional neural networks (CNN) [98, 20, 2, 65, 126]. Additionally, IR sensors mounted on the driver's seat headrest measure head position by detecting distance changes, aiding in fatigue detection [62].

Despite their effectiveness, vision-based methods suffer from severe limitations. Their performance degrades under poor lighting conditions, such as driving at night or in tunnels, and they become ineffective if the driver's face is partially or fully obscured by sunglasses, masks, or head movements. Furthermore, privacy concerns regarding continuous facial image monitoring limit their widespread adoption.

2.1.2 Wearable-based Approaches

Wearable sensors have been developed to monitor physiological signals associated with fatigue. Systems such as FEELythm, developed by Fujitsu [79], attach sensors to the driver's earlobe to monitor pulse rate variations. Other wearable devices, such as BioHarness chest belts [120], track respiration and heartbeat during driving, while EEG headsets analyze brainwave activity for early fatigue warnings [142]. Additionally, smart glasses equipped with proximity sensors can detect blink frequency as an indicator of drowsiness [13].

Although wearable sensors effectively capture physiological data, they introduce usability challenges. Prolonged wear can cause discomfort, reducing driver compliance. Moreover, most wearable systems focus on either physiological signals or facial fatigue cues—few solutions integrate both modalities, potentially affecting their robustness.

2.1.3 RF-based Approaches

RF-based fatigue detection methods exploit wireless signal reflections from the human body to infer fatigue-related behaviors. Research has explored WiFi [51], RFID [129], and radar-based [137] techniques for non-intrusive fatigue monitoring.

Long-wavelength RF signals, such as WiFi and RFID, primarily capture coarse-grained movements. For example, WiFind utilizes WiFi's Channel State Information (CSI) to detect nodding, yawning, and respiration activity [51]. Similarly, RFID tags attached to the driver's chest can track respiration rate fluctuations [129]. However, these signals have low spatial resolution due to their long wavelengths, making it difficult to distinguish fine-grained fatigue features across different body regions.

In contrast, short-wavelength, wide-band RF signals, such as UWB and mmWave radar, enable fine-grained fatigue feature extraction from multiple body regions. UWB radar-based solutions, such as V2iFi, extract subtle heartbeat rate variations from the driver [140], while Ubi-Fatigue detects both eye blinks and heartbeats using a 60 GHz radar [137]. However, UWB radars typically employ a single Tx-Rx pair, limiting their spatial sensing capabilities. This constraint makes it difficult to separate signals from different body parts—for example, distinguishing head and chest movements if they are in the same range bin.

mmWave radar overcomes these challenges by leveraging short wavelengths, wide bandwidth, and multiple Tx-Rx pairs. It enhances angular resolution, making it suitable for detecting both facial and physiological fatigue indicators simultaneously. Nevertheless, current mmWave-based approaches still have limitations. Some methods focus only on heartbeat detection [75], failing to capture the full spectrum of fatigue-related changes. Other studies target respiration and heartbeat analysis [30, 23] but overlook facial micro-movements, which are primary indicators of fatigue. Shu et al. [97] and Cardillo et al. [12] proposed eye blink detection techniques using mmWave radar, but their methods only identify individual blinks rather than fatigue patterns. Similarly, head motion tracking systems using mmWave radar [86, 95] focus on 3D head positioning but lack specific fatigue-related motion analysis.

2.1.4 Summary and Research Gaps

While vision-based and wearable-based fatigue detection methods offer useful insights, they suffer from real-world deployment challenges such as lighting conditions, user compliance, and privacy concerns. RF-based approaches, particularly mmWave radar, offer a promising alternative due to their non-intrusive nature, high resolution, and ability to simultaneously capture both facial and physiological fatigue indicators. However, existing mmWave-based methods remain limited in their ability to integrate multimodal fatigue cues from both head and chest regions.

This dissertation aims to address these limitations by developing a comprehensive fatigue detection framework leveraging mmWave radar. Compared to prior work, our system detects four facial and two physiological fatigue features, enabling a more robust and accurate assessment of driver fatigue in real-world driving scenarios.

2.2 Advancements in Acoustic-based Liquid Authentication

Recent research has mainly explored five techniques for liquid quality detection. Chemical and chromatographic techniques like GC-MS and HPLC detect contaminants in liquids such as water and milk but require bulky equipment and direct contact with the liquid [96, 33, 113, 46, 9, 69, 50, 115]. QET methods measure dielectric properties to detect flammable liquids but struggle with distinguishing real from fake liquids [29, 78, 101, 77]. Surface tension-based techniques, using tensiometers or cameras, identify liquids by measuring surface tension [109, 135, 47], but require direct contact and bulky equipment, making them impractical for sealed products. RF signal-based techniques analyze dielectric properties using RFID and UWB radar to identify liquids without opening containers [28, 40, 49, 27, 71, 41, 128, 117]. Acoustic techniques analyze sound wave interactions for non-contact detection [105, 118, 111, 133], with systems like Akte-Liquid [105] and HearLiquid [133] using low-cost microphones and speakers to detect liquid types. However, RF methods require non-commercial equipment, and both non-contact methods miss crucial chemical characteristics. In contrast, LiquidAuth employs a non-contact method providing molecular details, ensuring accurate authenticity detection.

GAN-based methods have become dominant in image recognition due to their ability to model complex high-dimensional distributions of images[16, 21, 37, 83]. Among them, image-to-image cGAN provides a mapping between two images[3, 119, 5, 1, 57, 68]. For example, Abhishek et al.[1] and Kim et al.[57] provide the transformation of medical images by synthesizing skin lesion images and transforming chest images into diagnostic images. For liquid quality detection, Li et al.[68] propose to enable the detection of the degree of contamination of water quality data by detecting temporal and spatial features in the water quality signals.

CNNs are widely used for image classification tasks[43, 91, 38, 6, 94, 93]. Some research focused on classifying the mass spectrum. For instance, Seddiki et al.[94] addresses the challenge of using CNN models for the classification of 1D mass spectra with small training sets, while Seddiki et al.[93] develops end-to-end CNN-LSTM models for early diagnosis using mass spectrometry data. In contrast, using FCNN, LiquidAuth achieved fine-grained classification of liquid category, brand, year, and authenticity.

2.3 Advancements in Adaptive Window Function Design for High-Resolution mmWave Radar Sensing

Robust mmWave radar sensing has evolved along three complementary research directions: adaptive window design for consistent performance across varying signal conditions, noise-aware angle estimation techniques resilient to hardware imperfections, and adaptive Doppler processing for reliable velocity estimation. This section reviews recent advances (2020–2024) in each area and discusses their limitations that motivate our proposed framework.

2.3.1 Adaptive Window Function Design

Window functions are fundamental components in radar signal processing that significantly impact system robustness by controlling the trade-off between mainlobe width and sidelobe suppression. Traditional fixed windows (Hamming, Chebyshev, etc.) offer predetermined trade-offs that cannot adapt to changing signal environments, limiting performance in dynamic scenarios with varying target characteristics and noise conditions. Recent research has explored more flexible and adaptive windowing approaches. Liu et al. formulated window design for synthetic aperture radar (SAR) imaging as a convex optimization problem, specifically a quadratically constrained quadratic program (QCQP), to jointly optimize the amplitude and phase of weighting coefficients [74]. Their method significantly reduced peak sidelobe levels in SAR range profiles compared to standard windows, enhancing performance in multi-target scenarios, albeit with some loss of signal-to-noise ratio (SNR). Cruz et al. conducted a comprehensive review of SAR image formation algorithms, including various window function designs and their impact on image quality [25]. Their analysis highlights the trade-offs between different windowing approaches and the need for adaptive techniques to optimize performance across varying operational conditions. Du et al. introduced optimization techniques for radar pulse compression using evolutionary algorithms [31]. These methods yield windows that outperform classical designs in sidelobe suppression while preserving resolution. While effective in specific environments, most approaches do not dynamically adapt to changing clutter or signal characteristics. Data-driven and real-time adaptive strategies have also emerged. Saeedi and Faez designed a nonlinear FM radar waveform whose spectrum implicitly shapes the windowing effect, combined with a mismatched filter to suppress sidelobes below –60 dB with minimal SNR loss [89]. Chowdhury et al. proposed a Peak-to-Sidelobe Ratio (PSLR) optimization scheme that dynamically adjusts window parameters based on current signal characteristics [22], demonstrating improved performance across varying SNR and clutter conditions. Despite these advances, most window design methods optimize the weighting function in isolation, without considering its impact on the entire processing chain, particularly angle estimation and Doppler processing. The designs often assume static or worst-case environments. Our PSLR-based adaptive windowing mechanism addresses this gap by dynamically optimizing parameters in real time, ensuring robust performance across changing operational conditions.

2.3.2 Noise-Aware Angle Estimation

Angle estimation is a critical component of radar sensing, and its robustness against non-ideal conditions has been extensively studied. Classical subspace-based algorithms such as MUSIC and ESPRIT remain popular for radar direction finding due to their high resolution under ideal conditions [35]. However, their performance degrades in practical scenarios with calibration errors, mutual coupling, and correlated noise. Li et al. proposed a Generalized Eigen-Decomposition (GED) version of MUSIC that explicitly models noise covariance, enabling accurate angle estimation in the presence of hardware imperfections [67]. Lu et al. developed a hybrid subspace-whitening approach that enhances DOA robustness under front-end distortions [76]. To tackle impulsive and non-Gaussian noise, Zheng et al. applied Bayesian sparse learning with Gaussian mixture modeling, achieving robust DOA estimation in challenging environments [139]. Rasekh et al. demonstrated that field-deployable calibration of phased arrays significantly improves angle estimation in real-world deployments [88]. Learning-based strategies have also emerged. Wu et al. introduced DiffRadar, a differentiable deep learning-based radar processing pipeline that improves angle estimation under hardware variability [125]. Chen et al. proposed mmHTSR, a transformer-based hybrid model integrating traditional signal processing with deep features, showing strong performance in low SNR [18]. While promising, these methods often require computational resources or detailed prior noise statistics. Our GED-MUSIC method strikes a balance between robustness and real-time feasibility by combining model-based noise adaptation with efficient iterative refinement.

2.3.3 Adaptive Doppler Processing

Doppler processing is essential for velocity estimation, particularly in automotive radar with dynamic multi-object scenes. Traditional FFT-based methods with fixed range bins struggle with range ambiguity and time-varying scattering. Han et al. reviewed Doppler processing techniques and highlighted the need for adaptability to handle urban clutter and variable dynamics [42]. Chen et al. proposed spatiotemporal Doppler tracking based on adaptive range-bin selection, achieving superior velocity accuracy in multi-target scenes [18]. Learning-based frameworks such as DiffRadar [125] adjust Doppler processing parameters in real-time but demand training data and tuning effort. Our spatiotemporal Doppler estimation framework bridges model-based control with adaptive tracking, maintaining accuracy in evolving environments. In summary, while prior methods make advances in individual areas, they often fail to integrate windowing, angle, and Doppler estimation in a unified pipeline. Our framework closes this gap with a joint, adaptive approach tailored for practical sensing.

Chapter 3

mmDrive: Fine-Grained Fatigue Driving Detection Using mmWave Radar

Abstract

Early detection of fatigue driving is pivotal for safety of drivers and pedestrians. Traditional approaches mainly employ cameras and wearable sensors to detect fatigue features, which are intrusive to drivers. Recent advances in radio frequency (RF) sensing enable non-intrusive fatigue feature detection from the signal reflected by driver's body. However, existing RF-based solutions only detect partial or coarse-grained fatigue features, which reduces the detection accuracy. To tackle above limitations, we propose a mmWave-based fatigue driving detection system, called mmDrive, which can detect multiple fine-grained fatigue features from different body parts. However, achieving accurate detection of various fatigue features during driving encounters practical challenges. Specifically, normal driving activities and driver's involuntary facial movements inevitably cause interference to fatigue features. Thus, we exploit unique geometric and behavioral characteristics of fatigue features and design effective signal processing methods to remove noises from fatigue-irrelevant activities. Based on the detected fatigue features, we further develop a fatigue determination algorithm to decide driver's fatigue state. Extensive experiment results from both simulated and real driving environments show that the average accuracy for detecting nodding and yawning features is about 96%, and the average errors for estimating eye blink, respiration, and heartbeat rates are around $2.21bpm$, $0.54bpm$, and $2.52bpm$, respectively. And the accuracy of the fatigue detection algorithm we proposed reached 97.63%.[53]

3.1 Introduction

Fatigue driving poses a serious threat to the safety of pedestrians and drivers and causes numerous economic losses, which have received extensive attention. Fatigue driving is one of the leading causes of traffic accidents, accounting for $20 - 40\%$ of total accidents [136]. In the USA, fatigue driving results in up to 6,000 fatal crashes each year [110]. In China, the number of fatal accidents caused by fatigue driving accounts for 80% of traffic accidents [134]. Furthermore, extensive research has shown that drivers exhibit typical fatigue features before the accident [85, 55]. Therefore, an effective fatigue driving detection method that can identify fatigue features and alert the driver in advance is highly required to reduce traffic accidents.

Driver fatigue features include driver's facial (e.g., nodding, yawning, and eye blink) and physiological (e.g., respiration and heartbeat) features. Traditional solutions mainly employ cameras and wearable sensors to detect facial [99, 7, 103, 32] and physiological features [54, 107], respectively. However, cameras mostly detect facial features and have poor performance under insufficient light or face-blocking (e.g., wearing masks and sunglasses) conditions [99, 7]. Meanwhile, drivers may be reluctant to reveal their facial images due to privacy concerns. Wearable sensors, such as EEG and ECG sensors mounted on the brain and chest, mainly measure the driver's physiological features and can invasively interfere with normal driving behavior [54, 107].

Compared with the abovementioned methods, recent advances in radio frequency (RF) sensing can enable light-independent and non-invasive fatigue detection. RF signal transmitter emits signal towards the driver. Then, the signal reflected from the driver's body, which involves the body movement information, is processed to detect fatigue features. However, existing methods that use long-wavelength RF signals, such as WiFi [52] and RFID [130], can only capture coarse-grained fatigue features (e.g., head nodding and yawing) but fail to obtain the tiny and fine-grained

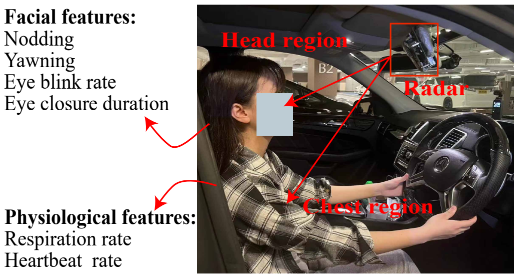


Figure 3.1: Deployment of mmDrive and fatigue features

features (e.g., eye blink or heartbeat). Besides, the single-transceiver design of RF devices (e.g., ultra-wideband radar [141]) only detects features from a single body area, i.e., either facial or physiological features around the chest.

However, relying on a subset of fatigue features may fail to achieve timely and accurate fatigue driving detection. This is due to the individual difference in manifesting feature fatigues [108]. For example, some people tend to show more facial fatigue features without changing their vital sign patterns when falling into the fatigue state. By contrast, others are more likely to experience more vital sign changes. In addition, the fine-grained features, especially the eye blink and eye closure duration, have been shown as essential alertness measures of fatigue driving [112]. Therefore, ignoring the fine-grained and various features from different body areas could hinder a holistic and precise detection of the driver's fatigue state.

In this paper, we fill this gap by developing a fatigue driving detection system using the millimeter-wave (mmWave) radar, called mmDrive, which can detect fine-grained facial and physiological features from different body areas. Our intuition is two-fold. First, the mmWave radar signal's wavelength is within several millimeters (e.g., 3.9mm for 77GHz radar), enabling a higher resolution to sense fine-grained fatigue features. Second, mmWave radars employ the frequency-modulated continuous-wave (FMCW) chirp signal, which can separate the signals from different body parts based on their distances to the radar. As shown in Figure 3.1, a compact mmWave radar is mounted in front of the driver sitting in the vehicle. The mmWave radar transmits the FMCW signal, which is reflected by the driver's body and propagates back to the radar. mmDrive can locate and separate the signals reflected from different body parts of the driver, i.e., head and chest regions. The amplitude and phase of signals from the head and chest regions are used to detect facial and physiological features, respectively. Finally, based on the detected fatigue features, we design a fatigue determination algorithm to decide whether the driver is fatigued or not.

Implementing the mmDrive system in practice, however, encounters several challenges. The first challenge stems from the driver's normal activities that could interfere with and confuse the detection of fatigue features. Specifically, apart from the head and chest movements in the fatigue state, drivers can move other body parts while driving, e.g., hand controlling the steering wheel. These disturbing body parts may move in a similar distance as the head or chest part to the radar. Then, the ranging ability of the FMCW chirp signal is unable to separate the interference brought by other body parts from the fatigue feature signals. In addition, many non-fatigue driver activities in the head region, i.e., head shaking and speaking, could introduce similar signal patterns as fatigue features, i.e., nodding and yawning. Therefore, these non-fatigue activities should be distinguished from fatigue features to guarantee detection accuracy.

To address this challenge, we leverage the unique geometric and behavioral characteristics of the fatigue features compared with other normal driver activities to achieve accurate fatigue detection. Specifically, we find that the interference from other disturbing body parts, which are located at the similar ranges, can be removed by adding the angle dimension. Thus, we first employ the beamforming technique to eliminate the interference from other body parts on the head and chest regions. Second, the fatigue features inherently exhibit distinctive patterns compared with the non-fatigue activities in the head region. For example, yawning is a slow-speed and low-frequency movement, while speaking involves more high-frequency variations in the signal. We can harness these differences to distinguish the fatigue features from other non-fatigue activities and movements.

The second challenge lies in the accurate extraction of the tiny eye blink feature. Although mmWave radar provides a higher resolution to sense fine-grained fatigue features, we face a practical issue for eye blink detection. Specifically, the eye blink feature suffers from the disturbance caused by involuntary facial movements, e.g., muscle twitching and arterial pulses on the face. As a result, the eye blink pattern in the measured signal phase involves much noise and fails to be clearly observed.

To tackle this issue, we propose to utilize the amplitude of the signal reflected from the facial region instead of the noisy signal phase to detect eye blinks. The rationale comes from the different reflection coefficients between the eyelid for closed eyes and the eyeball for opened eyes, which result in a conspicuous signal amplitude change during eye blink. By contrast, the signal amplitude caused by other inherent facial movements is relatively stable. Thus, the eye blink pattern can be notably observed from the signal amplitude.

In this paper, we propose a rule-based algorithm for detecting driver fatigue using multi-features. Our approach involves combining two different kinds of features: facial features and physiological signals. Our algorithm continuously detects physiological signals while also detecting facial features.

We select different facial features because different people exhibit different facial characteristics of fatigue. For example, some people may yawn when they are tired, while others may show increased eye closure duration. Therefore, we detect different facial features to account for individual differences. However, since facial fatigue features are highly correlated with a driver's fatigue state, one feature is sufficient for us to consider the driver as fatigued.

After detecting any facial feature, we examine the physiological features such as monitored heartbeat rate and breathing rate to determine whether they fall within the range of fatigue. If they do, we send a warning to the driver. If not, we continue detection in the next time window.

Even if no facial features are detected in a time window, we still determine whether the physical features meet the fatigue criteria because the driver's physiological features inevitably change when they feel tired. If they do meet the criteria, we send a warning to the driver.

In addition, we tried four other learning-based methods to determine driver fatigue. We trained four typical machine learning models using six driver features and fatigue labels as input, and compared the accuracy and F1-score of the results.

In sum, our paper makes the following contributions.

- To the best of our knowledge, we propose the first mmWave-based system for fine-grained fatigue driving detection, which can accurately detect various fatigue features from different body parts for determining the driver's fatigue state.
- We quantitatively model the effect of various fatigue features on the mmWave radar signal and dedicatedly design effective signal processing methods to eliminate the noises from other interfering activities and enable the accurate detection of tiny fatigue features.
- We conduct extensive experiments to evaluate the performance of mmDrive in both simulated and real driving conditions. The average accuracy for detecting nodding and yawning features is around 95%, and the average errors for estimating eye blink, respiration, and heartbeat rates are $2.21bpm$, $0.54bpm$, and $2.52bpm$, respectively. mmDrive can also detect facial features when the driver wears masks and glasses.

The remaining of this article is organized as follows. Section 3.2 discusses the characteristics of driver fatigue features and models the effect of different fatigue features on the mmWave radar signal. Section 3.3 presents the detailed design of the mmDrive system. Section 3.4 introduces the experimental settings and evaluation results of mmDrive. Section 3.5 introduces the related works for fagitue driving detection. Section 3.6 discusses the limitations of our work and suggests directions for future research. Section 3.7 concludes this article.

3.2 Fatigue Driving Features and Their Effects on mmWave Radar Signal

In this section, we first introduce the characteristics of driver fatigue features. Second, we model the effect of different fatigue features on the mmWave radar signal and demonstrate the underlying fatigue sensing principle.

3.2.1 Fatigue Driving Features

Existing driving safety research and studies show that driving fatigue is a transitional process with many early signs of the driver's behavior, which are called driver fatigue features [100]. There are mainly two categories of driven fatigue features, i.e., facial and physiological features. Facial features include nodding [32], yawning [7], eye blink [99], and eye closure duration [44]. The characteristics of facial features in the fatigue state are listed as follows.

- *Nodding*: When driving in the normal state, the driver looks ahead with the head positioned in a straight line with the body. However, when the driver falls into the fatigue state, the head could quickly tilt downward and then move upward. Specifically, the fatigue nodding feature is differentiated from the normal head nodding by the head pitch angle, i.e., the rotation angle θ of the head along the vertical direction. The fatigue nodding activity could incur the pitch angle to exceed 10°, while normal nodding has a smaller pitch angle [17].

- *Yawning*: When the driver feels fatigued, yawning is the body's attempt to keep alert and awake. During yawning, the mouth is opened to take a long and deep breath and then closed. The opening of the mouth usually lasts over $3s$ for the yawning activity in the fatigue state [132].
- *Eyes blink*: It has been established that the fatigue state can cause an increasing eye blink rate. In the normal state, the eye blink rate, which involves the movement of rapid closing and opening of the eyelid, is $10 - 15$ times per minute. While the eye blink rate can increase by $15 - 20\%$ which means jump to over 18 times per minute under the fatigue state [99].
- *Eye closure duration (ECD)*: ECD refers to the time when eyes are closed during a long eye blink. The fatigue will make the driver feel sleepy with an increasing ECD. Researchers have found that the ECD is greater than $1.2s$ for people under the fatigue state [59].

Physiological features include the driver's respiration [107] and heartbeat [54], which come from the movements in the chest region. During the fatigue state, the respiration rate and heartbeat rate both decline [54]. The normal respiration rate and heartbeat rate of adult drivers are within the range of $18 - 24$ breath per minute (bpm) and $60 - 100$ beats per minute (bpm), respectively. Whereas the fatigue state can make the driver's respiration rate reduce by $30 - 35\%$ which less than $12bpm$, as well as a slower heartbeat rate which reduce by $30 - 35\%$ which less than $60bpm$.

The above six fatigue features are complementary to each other for accurate fatigue driving detection. Extensive research studies have shown that a combination of various fatigue features can significantly improve the accuracy of fatigue driving detection [45, 15]. However, existing fatigue detection systems require the deployment of multiple intrusive sensors on the car and the driver's body, e.g., cameras and chest belts, to obtain the fatigue features from different body parts. By contrast, our work aims to realize accurate and non-intrusive fatigue driving detection only using a single mmWave radar.

3.2.2 mmWave Signal Primer

The mmWave radar emits the FMCW chirp signal $x(t)$, whose frequency linearly sweeps within a certain bandwidth: $f(t) = f_0 + kt$, where f_0 and k are the starting frequency and the chirp rate, respectively. The emitted signal $x(t)$ can be expressed as follows.

$$x(t) = e^{j2\pi f_0 t + j\pi k t^2}, \tag{3.1}$$

The signal sent to the driver will be reflected by different body parts of the driver and propagate back to the radar. Then, the received signal becomes:

$$y(t) = \sum_{i}^{N} \alpha_i \cdot x\left(t - \tau_i\right) \tag{3.2}$$

where α_i and τ_i refer to the attenuation factor and the time delay of the i^{th} signal path, respectively. The time delay τ_i is proportional to the propagation distance $d_i(t)$ of the signal path, i.e., $\tau_i = \frac{d_i(t)}{c} = \frac{2\left[d_{i_0} + \Delta d_i(t)\right]}{c}$, where d_{i_0} is the initial distance between the radar and the driver, $\Delta d_i(t)$ is the relative displacement of the driver body's movement, and c is the signal propagation speed.

Then, we obtain the intermediate frequency (IF) signal as $s(t) = x^*(t) * y(t)$, where $(\cdot)^*$ refers to the conjugate operation. The frequency components of the IF signal infer the range between the radar and different body parts. By performing the fast Fourier transformation (FFT) on the IF signal, which is called range-FFT, we can separately obtain the signal for different ranges $d_i(t)$ to the radar as below.

$$s_i(t) \approx \alpha_i e^{j4\pi f_0 \left[d_{i_0} + \Delta d_i(t)\right]} \tag{3.3}$$

Finally, we can extract the signal amplitude and phase of $s_i(t)$ from the corresponding

(a) Driver position

(b) Radar signal after Range-FFT

Figure 3.2: Driver, mmWave radar, and range-FFT result (signal amplitude, lighter area means higher amplitude)

range bin, which involves the driver's fatigue feature information from different body parts. In Figure 3.2, we depict the signal amplitude after range-FFT on the radar signal reflected by the driver in the cab, which exhibits two bright lines in different range bins. The front and back bins correspond to the head (d_{head}) and chest (d_{chest}) region, respectively, because the chest is farther from the radar than the head. In each range bin, the signal amplitude α_i measures the strength of the signal reflected from the body part. The signal phase ϕ reveals the fine-grained signal propagation distance: $\phi = 2\pi \cdot \frac{\Delta d_i(t)}{\lambda}$, where λ is the radar signal wavelength.

3.2.3 Modeling the Effect of Fatigue Features on the mmWave Radar Signal

To understand how the mmWave radar signal is utilized for fatigue feature detection, we perform quantitative modeling of the effect of each fatigue feature on the signal.

Nodding

The nodding feature under the fatigue state involves a quick head downward movement with a pitch angle (θ) over 10°, then followed by an upward head movement. As such, when the driver nods during fatigue, the distance between the head and radar will drop from d_{up} to d_{down} and then get back to d_{up}, as shown in Figure 3.3a.

(a) Nodding model

(b) Range-FFT result of nodding

Figure 3.3: Head movement and signal amplitude after range-FFT of the nodding feature

The distance change Δd of head nodding can be calculated based on the geometric relationship between the head position and pitch angle. Specifically, since the length of head and neck l_{hn} is $25-29cm$[114] for adults, $\Delta d = l_{hn} \cdot \tan\theta$ is within the range of $4.4-5cm$. On the other hand, the effective bandwidth B of the radar we use is $3.6GHz$, which corresponds to a range resolution of $d_{res} = \frac{c}{2B} = 4.17cm$. Thus, the nodding feature can cause the head range bin to shift one bin forward and then move backward, as depicted in Figure 3.3b. We can detect the nodding feature from the forward and backward shift of the head range bin.

Yawning

The yawning feature during fatigue mainly involves mouth movements, in which the mouth first gradually opens, stays opening for $3-7s$ to breathe in the air, and then closes to breathe out the air. The opening of the mouth will change the signal reflector from the outside lips to the soft palate inside the mouth, which increases the signal propagation distance from $2d_{mc}$ to $2d_{mo}$, as illustrated in Figure 3.4a. Then, closing the mouth will decrease the distance back to $2d_{mc}$. Since the distance $\Delta d_m \subset [2cm, 3cm]$ between the lip and soft palate is within the range resolution d_{res}, the signal phase of the head range bin can clearly show the yawning feature. As shown in Figure 3.4b, the signal phase first inclines when the mouth opens, then stays

(a) Yawning model

(b) Radar signal phase during yawning

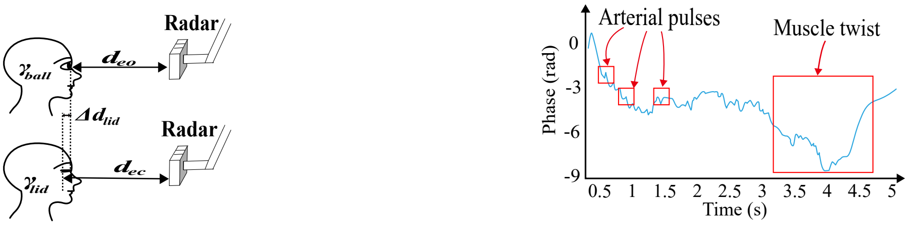
Figure 3.4: Mouth movement and signal phase during yawning

relatively stable for around $3s$, and finally declines with the closing mouth. Thus, we can detect the rising and falling edges and count the hold-up duration in the signal phase to detect the yawning.

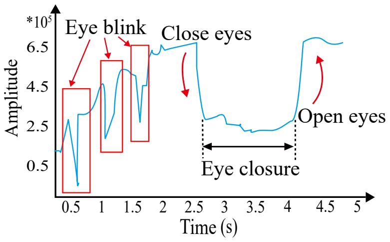
Eye Blink and Eye Closed Duration

Eye blink is a process in which the eyelids rapidly and repetitively close and open within a short period of time, i.e., $100-400ms$. When the eyelid is closed, the thickness of the eyelid introduces an extra signal propagation distance $2\Delta d_{lid}$ compared with the opened eyes, as shown in Figure 3.5a. Thus, we can intuitively employ the phase of the signal reflected from the eye, which is supposed to experience a sharp decline and incline process with a downward pulse pattern, to detect the eye blink. However, the measured signal phase from the head range bin in Figure 3.5b, does not exhibit the corresponding downward pulses when drivers blink their eyes three times in front of the radar. This is mainly because the involuntary muscle twist and arterial pulses over the whole face also incur a larger fluctuation of signal phase [90], which is superimposed with the signal phase caused by the tiny eyelid movement Δd_{lid} (¡1mm) [11]. Then, the eye blink pattern fails to be observed in the signal phase.

Therefore, we investigate the signal amplitude to extract the eye blink pattern. Our insight lies in the distinctive reflection coefficients of the eyeball (γ_{ball}) and eyelid



(a) Eye blink model (b) Noisy signal phase for 3 eye blinks



(c) Signal amplitude with 3 quick eye blinks and 1 long eye blink

Figure 3.5: Eyelid movement, signal phase, and signal amplitude during eye blink

(γ_{id}) due to their different materials. Through our experiments, we find that the eyeball has a higher reflection coefficient of the mmWave radar signal than the eyelid. This may be due to the reason that the surface of the eyeball is smoother than the eyelid skin. To this end, more signals reflected by the eyeball can directly propagate back to the radar instead of being scattered around by the uneven eyelid skin. As shown in Figure 3.5c, the measured signal amplitude with three eye blinks shows a clear decreasing pattern when closing the eye, followed by an increasing amplitude when opening the eye. Therefore, we can count the downward pulse in the signal amplitude as the number of eye blinks. Meanwhile, when the driver closes the eyes without opening them, the signal amplitude will decrease, stay at the lower level for a while, and finally increase to a higher level once the eyes are opened. Therefore, we can measure the time between the falling and rising edges of the signal amplitude as the eye closed duration (ECD).

(a) Respiration and heartbeat model

(b) Radar signal phase

Figure 3.6: Respiration and heartbeat movement and the corresponding signal phase

Respiration and heartbeat

When the driver breathes in and out the air, the chest will expand and contract accordingly. The chest movement during respiration brings a periodic change of the signal propagation distance within $2\Delta d_{\text{breath}}$, as depicted in Figure 3.6a. In addition, the heart bump introduces another distance change of $2\Delta d_{\text{heart}}$. These two movement effects are superimposed in the signal of the chest range bin with a sinusoidal change of the signal phase. Since $\Delta d_{\text{breath}} > \Delta d_{\text{heart}}$, respiration can result in a larger signal phase change than the heartbeat, as shown in Figure 3.6b. In addition, the heartbeat rate is higher than the respiration rate, which can be seen from the larger number of heartbeats than the respiration cycles. Hence, we can conduct frequency-domain analysis to extract respiration and heartbeat rates.

3.3 System Design of mmDrive

In this section, we introduce the detailed design of the mmDrive system. The overview of mmDrive is shown in Figure 3.7, which consists of four modules: (1) Signal preprocessing: detects the driver's movement status from the radar signal and reduces noise. (2) Body region separation: separate the signals reflected from different body regions. (3) Fatigue feature detection: detect the presence of nodding, yawning, eye blink, and estimate the eye blink rate, ECD, respiration rate, and heartbeat rate.

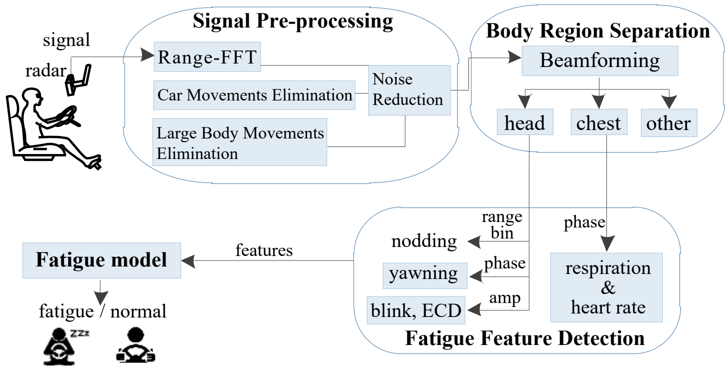


Figure 3.7: System overview

(4) Fatigue determination: input the detected features into a fatigue determination algorithm to decide whether the driver is in the fatigue state or not.

3.3.1 Signal Pre-processing

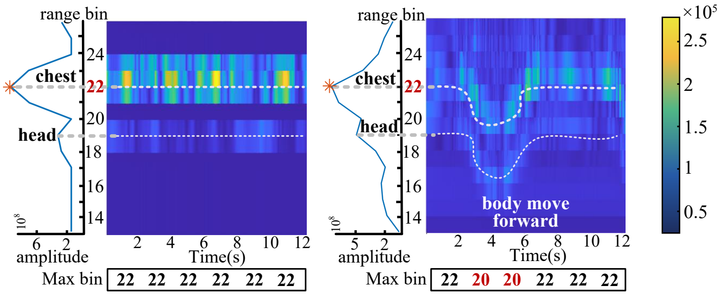
At first, we segment the incoming radar signal into fixed-length (i.e., $5s$) windows. Then, we perform range-FFT on each window's signal and obtain the signal amplitude for all range bins to detect the driver's chest and head. Since the large movements of cars and drivers will affect the received signal, before distinguishing the driver's chest area and head area, we need to pre-process the signal to reduce the impact of the large motion on feature extraction.

Driver-Passenger Spatial Discrimination While distinguishing between drivers and passengers presents inherent challenges in automotive radar sensing, our system addresses this issue through multiple complementary approaches. Under our mmWave radar configuration, the system achieves a spatial resolution of approximately 4 cm in the range domain, a performance metric obtained through the adoption of wideband frequency-modulated continuous wave (FMCW) radar signals, providing sufficient granularity to spatially separate occupants within the confined cabin environment. This resolution level enables the capture of minute displacement variations induced by human physiological activities, such as chest movements from respiratory patterns and subtle head motions.

This resolution enables reliable discrimination of driver and passenger positions across both azimuth and elevation angle domains through the construction of precise three-dimensional spatial positioning grids for target separation. The system employs multiple-input multiple-output (MIMO) antenna array technology, achieving angular resolutions of 2-3 degrees in both azimuth and elevation dimensions, ensuring the ability to distinguish targets at different spatial positions. Since the typical lateral separation between front-seat occupants (over 50 cm) far exceeds our system's spatial ambiguity threshold, this provides a substantial physical foundation for reliable occupant discrimination. According to automotive engineering standards, this spatial separation offers over 12 times the separation margin relative to the 4 cm range resolution, greatly exceeding the potential measurement errors of radar systems, thereby providing clear boundary conditions for the explicit definition of driver-specific monitoring regions and passenger areas.

Large Body Movements Elimination. In practice, the driver's body may experience large-scale movements due to sudden brake, vehicle vibration, or occasional body stretch. The radar signal can be significantly affected by these body movements. Hence, we also detect the windows with large-scale body movements in driver detection. To achieve this, we have two key observations. First, the driver's chest could move away from the original chest bin during large body movements. In other words, the chest range bin, which has the maximum amplitude, will change accordingly. As shown in Figure 3.8(b), the chest bin change from the 22^{th} to the 20^{th} bin during $2s - 6s$ when the driver's upper body moves forward. By contrast, when the driving is quasi-static, the chest bin keeps the same. To this end, we segment the whole window into short intervals (e.g., $0.5s$) and track the chest bin's index. A large body movement is detected if the index changes within the window.

Car Movements Elimination. Second, some body movements, e.g., those caused



(a) normal driving without large body movement

(b) body moving forward during driving

Figure 3.8: Signal amplitude of Range-FFT for driver detection

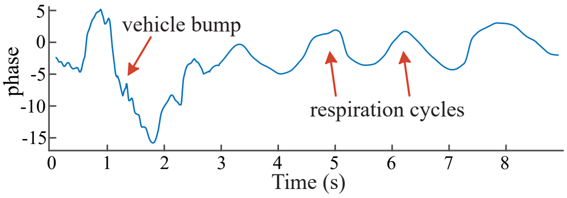


Figure 3.9: Signal phase of the chest bin when the vehicle encounters a deceleration zone

by sudden vehicle vibration, have smaller displacements and may not result in the shift of the chest range bin. While, these movements still incur the signal phase in the chest bin to experience more dramatic change than the tiny respiration and heartbeat activities. As depicted in Figure 3.9, the signal phase of the chest bin when the vehicle encounters a bump caused by the deceleration zone experiences a significant change than the respiration cycles. Thus, we also calculate the absolute difference between the maximum and minimum phase values in the window. If the phase difference is larger than $2 \cdot \frac{2\pi \Delta d_{breath_m}}{\lambda}$ (Δd_{breath_m} is the maximum breath displacement of $1cm$), we also regard the driver undergoes large body movement.

Noise Reduction. After removing the influence of large-scale movements, there will still be some other noises that can interfere with the detection of small fatigue features (eg, blinks, heartbeats). Noise will prevent the following feature extraction modules

(a) Original signal phase (b) Signal phase after noise reduction

Figure 3.10: Signal phase before and after noise reduction

to work properly because small eyes movements and heartbeat will be immersed in noise. Therefore, we use a FIR low-pass filter and a smoothing filter to increase the signal-to-noise ratio (SNR) of the signals. Considering that the frequencies of the six fatigue features are all below $2Hz$, we first set the stage frequency to $2Hz$ to filter the signal. The smooth filter with a window size of 50 points is used to further smooth the output signal of the low-pass filter. The signal comparison before and after filtering is shown in the Figure. 3.10. It can be seen that noise is suppressed.

3.3.2 Body Region Separation

Here we introduce the delay-and-sum (DAS) algorithm to help separate different body regions, such as the chest, head, and hand regions. By integrating the DAS algorithm with the virtual antenna technology, we can separate different regions located at similar ranges but at different angles.

Only relying on the range-FFT to explicitly separate the radar signals reflected from the chest and head regions is insufficient in real driving environments. This is because the driver's hand movement when controlling the steering wheel can interfere with the signal from the head region. Specifically, as illustrated in Figure 3.12a, the hand region has a similar distance as the head region to the radar, i.e., $d_{hand} \approx d_{head}$, leading to the superimposition of head and hand reflected signals in the similar range bins. Thus, we need to remove the effect from the hand region to detect facial fatigue features accurately.

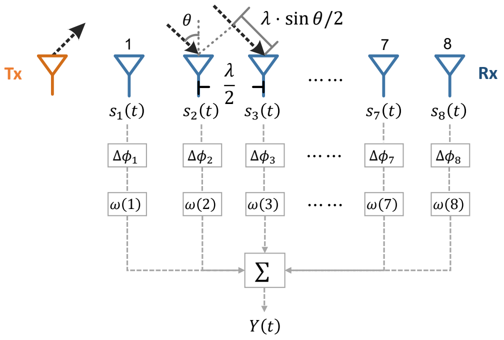


Figure 3.11: Delay-and-sum beamforming

.

To achieve this, we draw on the observation that the head and hand regions are located at different angles to the radar, i.e., θ_{head} and θ_{hand} in Figure 3.12a, to separate the signals from these two regions. In particular, we employ the multi-antenna design on the mmWave radar and adopt the beamforming technique for body region separation. We first rotate the radar by 90° to make its antenna array parallel to the vertical body line. As such, the first two transmitting antennas (Tx) and four receiving antennas (Rx) on the radar can formulate a virtual eight-element linear antenna array, as shown in Figure 3.11, to perform beamforming. Due to the distance interval $\frac{\lambda}{2}$ between adjacent antennas, a phase difference is introduced among the received signals across different antennas. The phase difference $\Delta\phi_m$ between the m^{th} Rx ($m \in [1, 2, ..., 7, 8]$) and the 1^{st} Rx is expressed as below.

$$\Delta\phi_m(\theta) = (m-1) \cdot \frac{2\pi}{\lambda} \cdot l_a \cdot \sin\theta, \tag{3.4}$$

where $l_a = \frac{\lambda}{2}$ is the distance interval between adjacent antennas. Then, we construct a signal steering vector $\omega(\theta)$ using $\Delta\phi_m(\theta)$ as follows.

$$\omega(\theta) = [e^{-j\Delta\phi_1(\theta)}, ..., e^{-j\Delta\phi_m(\theta)}, ..., e^{-j\Delta\phi_8(\theta)}] \tag{3.5}$$

Finally, the steering vector is multiplied with the received signal $s_{m,i}(t)$ from each antenna (m) and range bin (i) after range-FFT to obtain the beamforming signal $Y(i, \theta, t)$:

$$Y(i, \theta, t) = \sum_{m=1}^{8} \omega(m, \theta) \cdot s_{m,i}(t) \tag{3.6}$$

(a) Driving model with hand effect

(b) Signal amplitude after beamforming

Figure 3.12: Body region separation using beamforming

In Eq. (3.6), we traverse θ in the range of $[0, 180^\circ]$ with a step of 1°.

We apply the range-FFT and beamforming on the collected radar signals and show the beamforming signal amplitude in Figure 3.12b. Three highlighted clusters indicate greater signal amplitudes and correspond to the head, chest, and hand regions. The beamforming enables the separation of head and hand regions which are distributed on similar range bins. The experimental results are consistent with the theoretical prediction. The theoretical angular resolution of our mmWave radar is (assuming the target is located at approximately $\theta = 0$)

$$\theta = \frac{\lambda}{d \times N_{\text{Rx}} \times N_{\text{Tx}} \cos(\theta)} \approx 14.32^\circ. \tag{3.7}$$

And the angle difference between the head and chest region is more than 20°. In this way, we can exclusively extract the beamforming signal from the target head and chest regions for further fatigue feature detection. Note that hand region tracking through beamforming is also beneficial for detecting dangerous driving behaviors, such as hands off the steering wheel to pick up the phone.

3.3.3 Fatigue Feature Detection

In this section, we will utilize the range-FFT and beamforming to detect facial and physiological fatigue features.

In our scheme, we continuously detect the driver's location and six features for driver location initialization normal state feature data collection during the first five minutes after the driver starts the car. In general, drivers are typically not sleepy during the first five minutes of driving. Therefore, we can utilize the data from the first five minutes as the initialization data, assuming that the driver is in a non-fatigued state during this period. After the initial five minutes, we set the time window for feature detection to five seconds. This decision is based on our observation that changes in facial features associated with fatigue often occur within shorter durations, typically less than five seconds. Also, the average breathing cycle for humans is usually less than five seconds. However, since we have a narrow time window, there may be a problem with features being cut. Therefore, we choose to add a sliding window to prevent features from being cut and cannot be detected accurately. In the feature detection scheme, we have added the occurrence of the rising and falling edges of the signals that mark each feature (i.e., the marks of the appearance and end of the feature). If within a single time window, we find a rising signal edge of a feature but no falling edge, we continue to detect the feature by sliding the time window to the rising edge of the feature's signal. If the markers of rising and falling edges in a single time window can match, we detect the next time window.

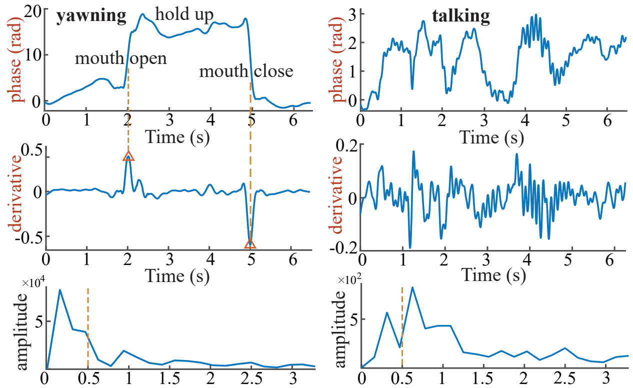
Nodding

Based on our modeling of the nodding feature on the radar signal, the head bin will shift forward and backward by at least one range bin caused by the head pitch angle over 10°. Therefore, we can track the head bin to detect nodding. Specifically, if the index of the head bin first decreases and then increases by 1-3 bins in the signal window without large body movements, we can determine the presence of fatigue nodding. Note that our method can exclude the condition by regarding the slight nodding conditions, e.g., the nodding when talking with other passengers, as the fatigue nodding. This is because the slight nodding has a smaller pitch angle less than 10° and does not change the head bin.

A practical concern in fatigue detection is the driver's head movements during driving, which can be mistaken for nodding. However, we found that head movements in other directions, such as leaning back, turning left, and turning right, increase the signal propagation distance between the radar and the head, while nodding shortens it. This allows us to distinguish nodding from other head movements. Regarding non-nodding head movements, we believe that when drivers tilt their heads left and right to look at the road, they are more conscious, and thus not in a state of fatigue. Similarly, when drivers are consciously talking or laughing, we can reasonably conclude that they are not fatigued.

Yawning

The mouth movements during yawning, i.e., mouth opening, holding up, and closing, incur the signal phase to increase, keep still, and decrease. Thus, we employ the phase of the beamforming signal from the head region to detect yawning. The first step is to detect the rising and falling edges, which correspond to the mouth opening and closing movements. An intuitive method is to detect whether the change in phase exceeds a specific threshold, but the shortcoming of this method is that it is difficult to set a threshold that meets different divers. And when the driver's head moves, the phase of yawning will be shifted a bit, then the detection method of threshold is no longer accurate. To avoid this, we design a *derivative-based* method inspired by the dramatic and unique change of the slope for the rising and falling edges. The derivatives for the rising edge undergo a sharp and high 'convex', while experiencing a



(a) Yawning: phase, derivative, FFT (b) Talking: phase, derivative, FFT

Figure 3.13: Yawning detection from the signal phase

sharp and high ‘concave’ for the falling edge, as shown in Figure 3.13(a). In addition, this method can effectively eliminate the effect of jitters in the signal phase because the derivatives of those smoother jitters are close to zero. As such, we can calculate the derivative of the signal phase and detect the presence of upper and lower peaks whose heights are over a predefined threshold σ_t[1]. We further calculate the time difference between the rising and falling edges to avoid mis-detecting inadvertent mouth open-close movements as the yawning. If the time duration is within $3-5s$, which is the common yawning duration for humans [81], we will regard the detected pattern as yawning.

A practical issue in yawning detection is that the driver’s other mouth-related activities, e.g., talking and laughing, can be mixed with the yawning feature. We think drivers are not fatigued when talking or laughing consciously. As depicted in Figure 3.13(b), the mouth movements during talking also result in fluctuations of the signal phase, going up and down. Hence, we need to differentiate yawning from other mouth activities. To address this issue, we leverage the observation that the talking and laughing activities involve more frequent fluctuations, as the underlying mouth movements can repeat the mouth open-close movement multiple times. Thereby, the

---

[1] The threshold σ_t is empirically set as the sum of the mean and two times of the standard deviation of all derivative values.

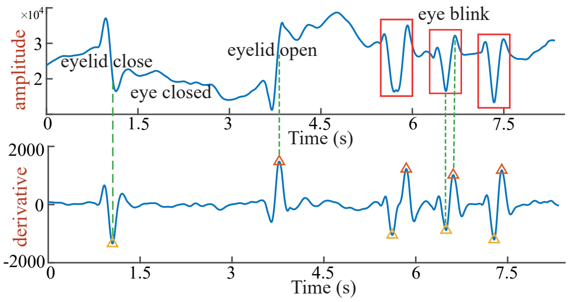


Figure 3.14: Eye blink detection and ECD measurement

corresponding derivatives of the signal phase experience higher jitters, resulting in a high-frequency component in the spectrum. Thus, to distinguish yawning from other mouth-related activities, we perform FFT on the derivative and compare the dominant frequency. As a result, the dominant frequency of yawning is less than $0.33Hz$ (the frequency of the shortest yawning duration of $3s$). By contrast, talking and other mouth-related activities have a higher dominant frequency over $0.5Hz$.

Eye Blink and Eye Closure Duration (ECD)

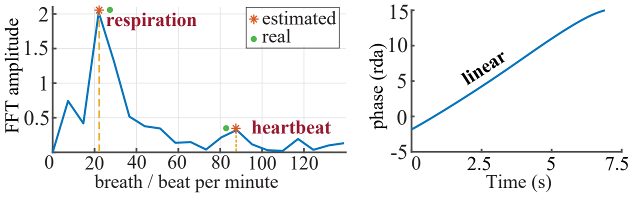
The eyelid first closes and then opens in an eye blink. Since the eyelid has a smaller reflection coefficient than the eyeball for the radar signal, the closure and opening of the eyelid cause the signal amplitude to first decrease and then increase. Accordingly, the quick eye blink incurs a local minima in the signal amplitude, and the longer eye closure leaves a relatively flat curve between the falling and rising edges of the eyelid closure and opening movements, as shown in Figure 3.14.

Thus, the *derivative-based* method is also applied to detect the eye blink and measure the ECD. First, the local minima of eye blink results in consecutive sharp ‘concave’ and ‘convex’ in the amplitude derivatives. Thus, we detect the presence of adjacent lower peak p_l and upper peak p_u whose heights are larger than σ_t. If p_l appears before p_u and the time difference t_{eye} between them is within the range of $100-400ms$, which is the normal blink period for humans, one normal eye blink is detected. As depicted in Figure 3.14, there are three sets of alternative p_l and p_u during $5.5 - 7.5s$ in the amplitude derivatives, which correspond to three eye blinks. The number of eye blink (n_e) in a window in T seconds is used to calculate the eye blink rate as $R_e = 60 \cdot \frac{n_e}{T}$ in the unit of blink per minute (bpm). Second, if t_{eye} between two adjacent p_l and p_u is longer than $500ms$, we will count it as a long eye blink and record down t_eye as the ECD. As illustrated in Figure 3.14, the first long eye blink has an ECD of around $2s$, indicating that the driver may fall into the fatigue state.

Respiration and Heartbeat

Both respiration and heartbeat incur the signal phase to change periodically. Meanwhile, respiration rate ($10-30bpm$) and heartbeat rate ($40-100bpm$) are distributed in different frequency ranges. Thus, we perform FFT on the signal phase to extract the respiration and heartbeat rates. For example, the FFT result of the phase in a window in Figure 3.15(a) exhibits a significant peak at around $22bpm$ (i.e., $0.366Hz \times 60s$) and a smaller peak at around $89bpm$ (i.e., $1.465 \times 60s$). As such, we can find the local peak in each specified frequency range to estimate the corresponding respiration and heartbeat rates.

However, simply applying FFT fails to obtain accurate estimation results due to the following reasons. First, the discrete FFT suffers from the frequency leakage problem, i.e., leakage of energy from the original frequency to its adjacent frequencies, which may overwhelm and hinder the detection of the smaller heartbeat peak. Second, the FFT resolution is the inverse of the window duration, i.e., $1/5s = 0.20Hz$, which leads to a coarse resolution of around $0.20Hz \times 60s = 12bpm$ for measuring respiration and heartbeat rates. As depicted in Figure 3.15(a), the estimated result from FFT fails to precisely match the real rate. Extending the window duration can improve the resolution. However, it sacrifices the time resolution to obtain the real-time fatigue state of the driver.



(a) FFT result of chest signal phase (b) phase of filtered signal

Figure 3.15: Respiration and heartbeat rate estimation

To tackle the above two issues, we apply a hamming window on the signal phase to overcome the frequency leakage problem. Next, we conduct two-step signal processing to acquire a fine-grained FFT result. First, we apply the zero-padding method before FFT by adding additional zeros to the end of the signal phase, which can identify the dominant respiration and heartbeat frequencies more accurately[2]. Second, we leverage the fact that a single-frequency signal has a linear phase of its complex time-domain signal, which can be used to accurately estimate the frequency [4]. Thus, after zero-padding and FFT, we only keep the peak frequency and its two adjacent frequencies of the respiration and heartbeat in the FFT result, respectively, so that the effect from other frequencies is removed. Then, we conduct inverse FFT (IFFT) on the filtered FFT result and attain a complex time-domain signal. We calculate the slope of the linear signal phase after IFFT as $\hat{m}$, as shown in Figure 3.15(b). Finally, we obtain the respiration (R_r) and heartbeat rate (R_h) as $R_{r/h} = 60 \cdot \frac{\hat{m}_{r/h}}{2\pi}$ in the unit of bpm.

3.3.4 Fatigue Determination

We use the method in Section 3.3 to obtain data on six features of the driver, including nodding, yawning, eye blink, ECD, respiration, and heartbeat rate. In this section, we propose a rule-based fatigue determination algorithm and select four learning-based methods for comparison based on the six features we extracted before. In the rule-based fatigue determination algorithm, we define the detection priority of

---

[2] The total length of phase is doubled after zero-padding.

facial features and simultaneously use facial features and physiological features to determine the driver's state. In the learning-based fatigue detection model, we trained four machine learning models: Support vector machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), and fully connected neural network (NN).

Rule-based fatigue determination algorithm

We took the driver's state during the first five minutes of driving as the normal state for reference and driver region detection. After the first five minutes, we set the time windows to be every five seconds.

According to the above observations, we design the fatigue determination algorithm as illustrated in Algorithm 1. First, we input the radar signals within a time window after range-FFT and beamforming. Then, we detect whether there is large-scale body movement by the driver. If so, we wait for the next signal. If not, we will simultaneously process the signals from the head and chest region according to priority, and then compare them with the data in the normal state to determine whether the driver is fatigued.

In the same kind of features, we still choose different features because different people show different facial features of fatigue (for example, some people will yawn when they are fatigue, while some people will not yawn, but their eye closure duration will increase), so we will detect different features to account for individual differences. However, since the facial fatigue feature is strongly correlated with the driver's fatigue state, one feature is sufficient for us to consider the driver to be fatigue. To prevent us from mistaking the movement of other head areas of the driver as a sign of fatigue, after we detect any facial feature, We will continue to judge whether the driver's physiological signals are outside the normal range. Meanwhile, when there is no facial fatigue feature, we will directly determine whether the driver is fatigue or not according to the driver's physiological signal. The specific steps are as follows:

Algorithm 1 Fatigue determination from driver fatigue features in one time window

```
Input: Signal in a window ($w$)
Output: Fatigue decision $D_f$ ($D_f = 0$: not fatigue, $D_f = 1$: fatigue)
if Driver body is moving then
    return $D_f = 0$
end if
if Driver's head is moving then
    if Nodding is true then
        $D_f \leftarrow 1$
    end if
else if Driver's mouth is moving then
    if Yawning is true then
        $D_f \leftarrow 1$
    end if
else if Eye blink rate increase $\geq 15\%$ or ECD $\geq 1.2$ s then
    $D_f \leftarrow 1$
end if
if Respiration rate decrease $\geq 30\%$ then
    $D_f \leftarrow 1$
end if
if Heartbeat rate decrease $\geq 30\%$ then
    $D_f \leftarrow 1$
end if
if $D_f = 0$ then
    No alert
else if $D_f = 1$ then
    Send alert
end if
```

Considering that different facial features can affect each other, for example, a large-scale mouth movement when yawning can affect the detection of eye blinking. Therefore, facial features are detected in priority order of head ¿ mouth ¿ eyes. First, we use the head bin's signal phase to detect whether the driver's head is moving. To achieve this, we calculate the absolute difference between the maximum and minimum phases in the window. Because head movements result in a more significant signal propagation distance than mouth movements, if the phase difference exceeds $2 \cdot \frac{2\pi\Delta d_m}{\lambda}$ (d_m is the upper bound of the distance between human lips and the soft palate, about $4cm$ [60]), we will detect the head is moving.

Then, we track the head range bin to detect nodding. If not, we will continue to use the head bin's signal phase to detect mouth movements. As the movement of the eyelid during blinking is smaller than that of the mouth, if the absolute phase difference in the window is greater than $2 \cdot \frac{2\pi\Delta d_{lid}}{\lambda}$ (d_{lid} is the thickness of human eyelids, about $1mm$ [73]), we consider that the mouth is moving, After confirming the mouth is moving, we detect the yawning. But if there is no mouth movement, the eye blink and ECD will be detected and measured. When the number of eye blink exceeds ten times per minute, or the ECD is over $1.2s$, the driver is considered fatigued.

At the same time, we use the chest area signal in the window to measure physiological features. If the breathing rate or heartbeat rate decreases by more than 30% of the normal state, the driver is considered fatigued.

Combining these steps, the algorithm issues a weak warning to the driver when one of the physiological and facial features is detected as fatigue. If both features are detected as fatigued, a strong warning is issued to remind the driver to rest.

Learning-based fatigue determination model

In this subsection, we introduce the machine learning model we use and the data collection method.

Data collection: We selected 20 volunteers (10 males, 10 females) and collect data under two conditions. In the first condition, the volunteers were asked to drive in a simulation environment for 3 hours. Volunteers actively reported when they felt fatigue, and we selected the data after the report time and the data five minutes before the report time (60 time windows) to mark as fatigue data. In the second condition, the volunteers started to drive for half an hour in the simulation environment when they reported feeling fatigue. The reason for designing these two data collection methods is that in the first condition, there are more features of the driver's normal driving state and fewer features of fatigue. In order to balance the amount of data used to train the model, we use the second condition to increase the amount of data with fatigue features. We extract the features contained in each time window (5 seconds) into the following array:

$$F_i = \left(nodding_i, yawning_i, blinkrate_i, ECD_i, breath_i, heartbeat_i\right), \tag{3.8}$$

where i represents the index of the time window, $nodding_i$ represents the number of nodding, $yawning_i$ represents the number of yawning, $blinkrate_i$ represents the eye blink rate with the unit of bpm, ECD_i represents the eye closure duration with the unit of millisecond, and $breath_i$ represents the breathing rate with the unit of bpm, $heartbeat_i$ means heartbeat rate with the unit of bpm. We set the data label of the normal state to 0 and the label of the fatigue state to 1. We finally selected 22,000 time windows for both the fatigue state and the normal state to form a training set with 40,000 time windows and a validation set with 4,000 time windows. So the dataset used for training is represented as:

$$Dataset = \left(F_i, label_i\right), \tag{3.9}$$

where $i = 1, .., 44000$.

In order to better compare which method is more advantageous in solving the problem of individual differences, we also designed a second experimental method. We select the data of 20 volunteers (5 males and 5 females) with a total of 20,000 time windows (10,000 fatigue states and 10,000 normal states) as the training set, and then we select 4,000 time windows data from the remaining ten volunteers (2,000 time windows of fatigue states and 2,000 time windows of normal states) used as a validation set to detect whether the model can better adapt to individual differences. Next, we introduce the four machine-learning models we used.

Support vector machine (SVM): SVM is a classification algorithm separating data items. This algorithm was proposed by Vladimir N. Vapnik based on statistical learning theory [24]. The main purpose of the SVM is to find the best hyperplane to distinguish the data given as two-class or multi-class. The method we used was carried out in two classes. Whereas label 1 means that the driver is fatigue, label 0 means the driver is in a normal state. We aimed to distinguish fatigue drivers from non-tired drivers.

The kernel function we choose is gaussian kernel function, the decision function is one vs rest, and the error accuracy of stopping training is 0.001.

Naive bayes: Based on our labeled data, we created a Naive Bayesian classifier for driver fatigue status determination. Our model consists of (class) variables C, including two mutually exclusive states c_1, c_2, i.e. normal state and fatigue state, and a set of attributes $A1, A_2, A_3, A_4, A_5, A_6$ (i.e. number of nodding, number of yawning, eye blink rate, eye closure duration, respiration rate, and heartbeat rate). The goal of this layer is to find the probability of class state c_i:

$$\operatorname{Pr}\left(C=c_i \mid A_1=a_1, A_2=a_2, A_3=a_3, A_4=a_4, A_5=a_5, A_6=a_6\right) \tag{3.10}$$

when this probability is maximum, the state of the driver belongs to the class variable c_i.

We chose Naive Bayes with Gaussian distribution prior (GaussianNB). GaussianNB assumes that the prior probability of the feature is a normal distribution, which is as follows:

$$P\left(X_j=x_j \mid Y=C_k\right)=\frac{1}{\sqrt{2 \pi \sigma_k^2}} \exp \left(-\frac{\left(x_j-\mu_k\right)^2}{2 \sigma_k^2}\right) \tag{3.11}$$

where C_k is the k^{th} category of Y. μ_k and σ_k^2 are the values that need to be estimated from the training set. GaussianNB will find μ_k and σ_k^2 based on the training set. μ_k is the mean of all X_j in sample category C_k. σ_k^2 is the variance of all X_j in sample category C_k.

There is only one main parameter of the GaussianNB class we used, namely the prior probability, corresponding to the prior probability $P\left(Y=C_k\right)$ of each category of Y. We didn't give the prior probability in our methods, therefore prior probability is $P\left(Y=C_k\right)=m_k / m_.$ where m is the total number of training set samples, and m_k is the number of training set samples whose output is the k category.

K-Nearest Neighbor (KNN): The KNN method classifies objects based on the learning data closest to them. Learning data will be projected into multiple dimensional spaces, each dimension representing a feature of the data. In our model, the space is classified into two parts according to the learning data classification, namely normal state and fatigue state. If the focal class is the most common classification of KNearest Neighbor at that moment, then a point in that space is labeled as the focal class. Nearest or distant neighbors are usually calculated based on Euclidean distance:

$$where, \operatorname{dist}(X, Y)=\sqrt{\sum_{i=1}^n \quad \left(X_i-Y_i\right)^2}$$

$$\begin{aligned} X_i &: \text{The training data} \\ Y_i &: \text{The testing data} \\ n &: \text{The number of training data} \end{aligned} \tag{3.12}$$

In our model, we choose the number of neighbors to be 15.

Neural network (NN): The training of the neural network is divided into two stages: forward propagation and back propagation. The forward propagation process of the network first puts the feature array into the input layer, followed by further feature extraction through the weighted summation of the input and the network weight through the 5-layer fully connected layer. Finally, the softmax activation function is applied to obtain the probability values for classification into different driver states, with the highest probability indicating the predicted driver state. The network weights are trained by the back propagation algorithm, and the loss function uses the square error. The gradient descent method is used to update the weights, with the gradient of the weights calculated based on the loss function. The weights are then iteratively updated in the opposite direction of the gradient in order to minimize the loss function.

3.4 Evaluation

In this section, we introduce the experimental settings and evaluation results of mmDrive.

3.4.1 Experiment Setup and Evaluation Method

Implementation and evaluation environments

The mmDrive system is implemented using a commodity mmWave radar, Texas Instruments (TI) AWR1843BOOST, working on the $77-81GHz$ frequency band with

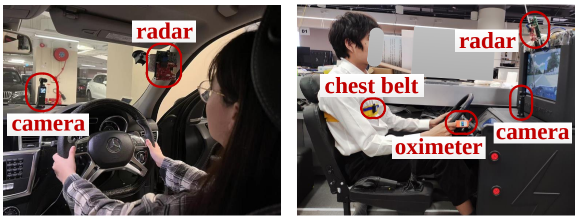


Figure 3.16: Experiment setup and driving environment

three Tx and four Rx antennas. The radar signal is collected in real-time using the TI DCA1000EVM board with a sampling rate of $2KHz$. The signal is processed in Matlab on a laptop with an Intel i7-1165G7 processor and 32GB memory.

We conduct experiments in both simulated and real driving environments, as shown in Figure 3.16. The radar is installed above the user and below the windshield of a vehicle.

We find 20 volunteers, including 10 males and 10 females, to test our system. In the simulated environment, they drive on virtual urban roadways for over three hours. In the real driving environment, volunteers are asked to drive in various road conditions, including the residential area, the highway, and a parking lot. In the residential area, the vehicle is driven under around 40-50 miles per hour. There are around 6 crossroads during driving. On the highway, volunteers drive at the speed of 90-100 miles per hour. In the parking lot, the vehicle is driven at a lower speed of around 20-30 miles per hour. Under each road condition, we find two volunteers to drive for around half an hour. Another two volunteers are sitting on the copilot and back seats. To measure the ground truth of fatigue features, we place a DJI Pocket 2 camera in front of the drivers to capture their facial videos during driving as the ground truth for the nodding, yawning, and eye blink features. Then, we carefully look through the video to mark the presence of facial fatigue features and match its time frame with the radar signal. The ground truth of respiration and heartbeat rates are collected from the accelerometer-based chest belt and oximeter, respectively.

Experimental Dataset Diversity

Although our experimental dataset comprises a relatively modest scale in absolute terms ($\sim$20 volunteers contributing approximately 200 hours of driving data), we implemented a systematic participant recruitment strategy designed to maximize demographic and physiological representativeness within the available sample size. The volunteer cohort was deliberately stratified across multiple dimensions to ensure comprehensive coverage of real-world driver variability, recognizing that dataset diversity can partially compensate for limited sample size in establishing proof-of-concept validity.

The volunteers included both male and female drivers with balanced gender distribution (50% male, 50% female), spanning a broad age range from 21 to 55 years to capture generational differences in driving behavior and physiological characteristics. Significant anthropometric variation was incorporated, with participants exhibiting varying eye sizes (ranging from 2.3 to 3.1 cm in horizontal diameter), head circumferences (52 to 61 cm), and overall body heights (155 to 185 cm). This morphological diversity is particularly critical for mmWave-based physiological sensing, as radar cross-sectional properties and signal reflection patterns exhibit substantial inter-individual variation based on tissue density, facial bone structure, and overall head geometry. Additionally, participants represented different ethnic backgrounds, contributing to natural variation in skin tone, facial structure, and hair characteristics that could influence radar signal propagation and reflection.

In addition to demographic stratification, physiological variations such as different baseline respiration rates (12–20 breaths per minute) and heart rates (58–89 beats per minute) were naturally included across the participant pool. Some participants exhibited faster cardiovascular rhythms associated with higher metabolic rates or caffeine consumption, while others demonstrated slower, more regular patterns typical of individuals with higher cardiovascular fitness or meditation practices. This heterogeneity in fundamental physiological signatures provides essential diversity in the temporal patterns that our fatigue detection algorithms must accommodate. Furthermore, participants varied in their sleep quality patterns, circadian preferences (morning vs. evening chronotypes), and baseline fatigue susceptibility, contributing to natural variation in fatigue onset characteristics and manifestation patterns during experimental sessions. This multifaceted diversity provides broader coverage of real-world driver variability, partially compensating for the limited dataset size while establishing a robust foundation for algorithm development and initial validation.

Evaluation metrics

First, for the task of driver detection and each feature detection, we have used three evaluation metrics namely accuracy, specificity, and sensitivity. The formulae for calculating these values are given as under:

$$Accuracy_{feature} = \frac{TP_{feature} + TN_{feature}}{TP_{feature} + TN_{feature} + FP_{feature} + FN_{feature}} \tag{3.13}$$

$$Specificity_{feature} = \frac{TN_{feature}}{TN_{feature} + FP_{feature}} \tag{3.14}$$

$$Sensitivity_{feature} = \frac{TP_{feature}}{TP_{feature} + FN_{feature}} \tag{3.15}$$

where the value of $feature$ is one of the head, chest, ds (driver moving state), nodding, and yawning.

For driver moving state detection, TP_{ds} (True Positives) are those cases where a driver moving state is classified as static state and TN_{ds} (True Negatives) are those when a driver static state is classified as a static state. Similarly, FP_{ds} (False Positives) corresponds to the cases where a driver static state is classified as a moving state, and FN_{ds} (False Negatives) to the cases where a driver moving state is classified as a static state.

For head region detection, TP_{head} are those cases where a head region is classified as head region and TN_{head} are those when a non-head region is classified as a non-head region. Similarly, FP_{head} corresponds to the cases where a non-head region is classified as a head region, and FN_{head} to the cases where a head region is classified as a non-head region.

For chest region detection, TP_{chest} are those cases where a chest region is classified as chest region and TN_{chest} are those when a non-chest region is classified as a non-chest region. Similarly, FP_{chest} corresponds to the cases where a non-chest region is classified as a chest region, and FN_{chest} to the cases where a chest region is classified as a non-chest region.

For nodding detection, $TP_{nodding}$ are those cases where a nodding is classified as nodding and $TN_{nodding}$ are those where a non-nodding head movement is classified as non-nodding. Similarly, $FP_{nodding}$ corresponds to the cases where a non-nodding head movement is classified as nodding, and $FN_{nodding}$ to the cases where a nodding is classified as non-nodding.

For yawning detection, $TP_{yawning}$ are those cases where yawning is classified as yawning, and $TN_{yawning}$ are those when a non-yawning mouth movement is classified as yawning. Similarly, $FP_{yawning}$ corresponds to the cases where a non-yawning mouth movement is classified as yawning, and $FN_{yawning}$ to the cases where yawning is classified as non-yawning mouth movement.

Second, we calculate the mean absolute error (MAE) of eye blink rate, respiration rate, and heartbeat rate, as $MAE = \frac{\sum_{i=1}^{n} r_i - r_i'}{n}$, where n is the total number of windows, and r_i and r_i' refer to the real and estimated rates. Finally, the average error (e_{ECD}) between the real and measured ECD from the video and radar is obtained.

Third, for the fatigue determination algorithm and model, we use Accuracy and F1-score as the metrics and the definition of accuracy and F1-score are as follows:

$$
\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}_{\text{fatigue}} + \text{TN}_{\text{fatigue}}}{\text{TP}_{\text{fatigue}} + \text{TN}_{\text{fatigue}} + \text{FP}_{\text{fatigue}} + \text{FN}_{\text{fatigue}}} \\
\text{Recall} &= \frac{\text{TP}_{\text{fatigue}}}{\text{TP}_{\text{fatigue}} + \text{FN}_{\text{fatigue}}} \\
\text{Precision} &= \frac{\text{TP}_{\text{fatigue}}}{\text{TP}_{\text{fatigue}} + \text{FP}_{\text{fatigue}}} \\
\text{F1-score} &= \frac{2 \cdot \text{TP}_{\text{fatigue}}}{2 \cdot \text{TP}_{\text{fatigue}} + \text{FN}_{\text{fatigue}} + \text{FP}_{\text{fatigue}}} = \frac{2 \cdot \text{ Precision } \cdot \text{ Recall}}{\text{Precision } + \text{ Recall}},
\end{aligned}
\tag{3.16}
$$

where $TP_{fatigue}$ are those cases where a fatigue state is classified as fatigue and $TN_{fatigue}$ are those when a normal state is classified as fatigue. Similarly, $FP_{fatigue}$ corresponds to the cases where a normal state is classified as fatigue, and $FN_{fatigue}$ to the cases where a fatigue state is classified as a normal state.

We found that recall reflects the ability of the model and algorithm to identify positive samples. The higher the recall, the stronger the model's ability to identify positive samples. The precision reflects the ability of the model to distinguish negative samples. F1-score is a combination of both. The higher the F1-score, the more robust the classification model.

3.4.2 Evaluation Results

Performance of driver detection

We first evaluate the performance of our method to detect the driver's state and the target head and chest regions. We randomly select 1000 signal windows (5s for each window) from the collected radar signal in which all volunteers are either in the normal quasi-static driving state (500 time windows) or the body moving state (500 time windows). Meanwhile, we also evaluate the accuracy, specificity and sensitivity of the head and chest region detection when the volunteer is quasi-static. The evaluation results are given in Table 3.1 and Table 3.2. The accuracy of driver movement state detection in simulated and real driving environments are both exceed 93%. The chest region detection can be achieved with higher accuracy because the chest region has a larger area than the head region to reflect the radar signal.

Performance of nodding detection

Next, we evaluate the performance of detecting each fatigue feature. We randomly select 50 time windows from the collected radar signal for each volunteer in which the volunteer is in the normal quasi-static driving state, and half of the windows have the nodding feature (i.e., 25 windows with the nodding feature). We selected a total of 2000 time windows. Meanwhile, we also evaluate the performance of our method to avoid detecting other irrelevant head movements as nodding. Volunteers are asked to randomly shake their heads up, left, and right. They also mimic the slight nodding during talking. Then, we calculate the accuracy, specificity, and sensitivity of the nodding feature. As depicted in Table 3.3, the $accuracy_{nodding}$ is within $95.7\% - 99.6\%$ with an average of 98.30% in the simulated driving environment and within $94.75\% - 98.3\%$ with an average of 96.85% in the real driving environment. The results show that our nodding detection method can effectively detect and distinguish the nodding feature from other head movements.

Performance of yawning detection

We randomly select 50 signal windows from the collected radar signal for each volunteer in which volunteer is in the normal quasi-static driving state and half of the time windows have the nodding feature (i.e., 25 time windows with the nodding feature). We selected a total of 2,000 time windows.

We evaluate the performance of yawning detection. The accuracy of yawning detection for each volunteer is shown in Figure 3.17. The $accuracy_{yawning}$ is obtained by dividing the number of accurately detected yawning over the total number of yawning.

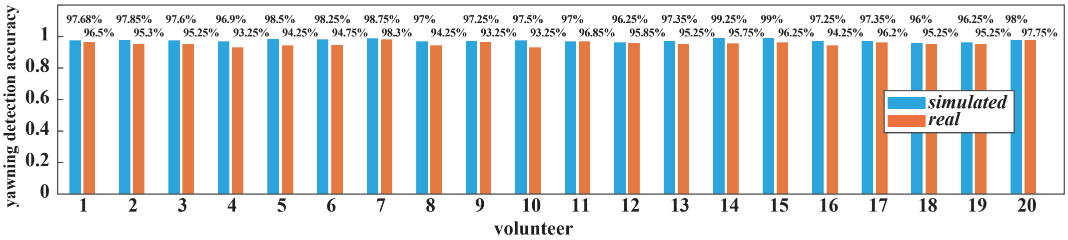


Figure 3.17: Performance of yawning detection

Table 3.1: Results of driver moving state and region detection in simulated environments

Item	TP	TN	FP	FN	Accuracy	Specificity	Sensitivity
ds	489	476	24	11	96.50%	95.20%	97.80%
head	481	497	3	19	97.80%	99.40%	96.20%
chest	490	492	8	10	98.20%	98.40%	98.00%

The $accuracy_{yawning}$ ranges from 96% to 99.25% with an average of 97.55% among all volunteers in the simulated driving environment and ranges from 93.25% to 98.3% with an average of 95.50% in the real driving environment.

In addition, we also evaluate the robustness of our yawning detection method in the face of other mouth-related movements, such as talking, laughing, and coughing. For example, during driving, we ask other passengers in the vehicle to chat with the driver and let the driver deliberately cough several times. Therefore, we calculate the specificity and sensitivity of the yawning feature. As depicted in Table 3.4, the average sensitivity of all volunteers is 96.7% and 94.6% in simulated and real driving environments, respectively. And the specificity is over 95% in both simulated and real driving environments, which ensures the robustness of our yawning detection method.

Table 3.2: Results of driver moving state and region detection in real environments

Item	TP	TN	FP	FN	Accuracy	Specificity	Sensitivity
ds	477	462	38	23	93.90%	92.40%	98.40%
head	473	485	15	27	95.80%	97.00%	94.60%
chest	486	488	12	14	97.40%	97.60%	97.20%

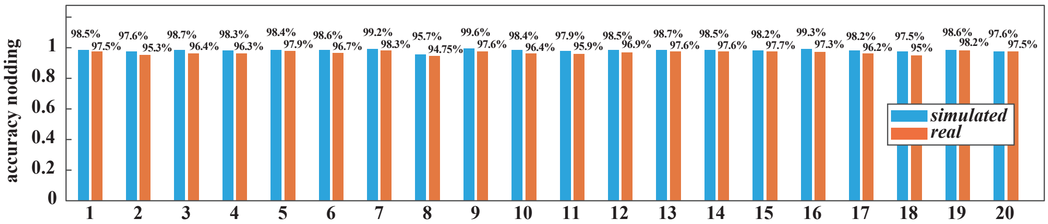


Figure 3.18: Performance of nodding detection

Performance of eye blink rate and ECD estimation

Third, we evaluate the performance of estimating the eye blink rate and ECD. We calculate the MAE between the real and estimated blink rates for each volunteer in both simulated and real driving environment, as illustrated in Figure 3.20. The MAE of eye blink rate are around $1-3$ and $1.75-3.5$ blink per minute in simulated and real driving environment respectively. The average MAE among different volunteers are 1.74 and 2.68 blink per minute in simulated and real driving environment respectively. We find that the MAEs of volunteer 3 and 8 are relatively higher than others. This is mainly because their eye blink habit is different. They tend to blink fast but slightly with two to three continuous blinks together, during which our method may miss one of the blinks. Then, we calculate the error in ECD measurement (e_{ECD}) for long-duration eye closure in both simulated and real driving environment, as shown in Figure 3.21. The e_{ECD} ranges within $15-45ms$ and $20-45ms$ with an average of $31.5ms$ and $33ms$ in simulated and real driving environment respectively. Compared with the ECD of over $1.2s$ during the fatigue state, the e_{ECD} is quite small, indicating

Table 3.3: Results of nodding detection

Environments	TP	TN	FP	FN	Accuracy	Specificity	Sensitivity
Simulated Environments	979	987	13	21	98.30%	98.70%	97.90%
Real Environments	962	975	25	38	96.85%	97.50%	96.20%

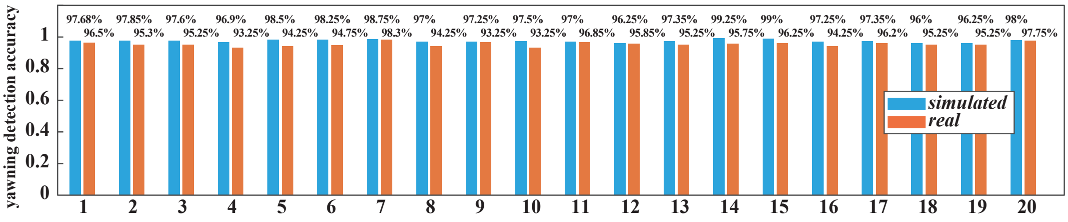


Figure 3.19: Performance of yawning detection

a high accuracy for ECD estimation.

Performance of respiration and heartbeat rate estimation

Finally, we evaluate the performance of estimating the respiration rate and heartbeat rate. The MAEs between the real and estimated respiration and heartbeat rates are shown in Figure 3.22 and Figure 3.22, respectively. The MAE of respiration rate (range: $0.15 - 1$ breath per minute, average: 0.54 breath per minute) estimation is lower than that of the heartbeat rate (range: $0.5 - 5$ beat per minute, average: 2.52 beat per minute). This is mainly because the chest displacement during respiration is larger than the tiny pulse of heartbeat. Nevertheless, considering the heartbeat rate range at around $60bpm$, the average MAE of $2.52bpm$ is in a decent error level.

Table 3.4: Results of yawning detection

Environments	TP	TN	FP	FN	Accuracy	Specificity	Sensitivity
Simulated Environments	967	984	16	33	97.55%	98.40%	96.70%
Real Environments	946	964	36	54	95.50%	96.40%	94.60%

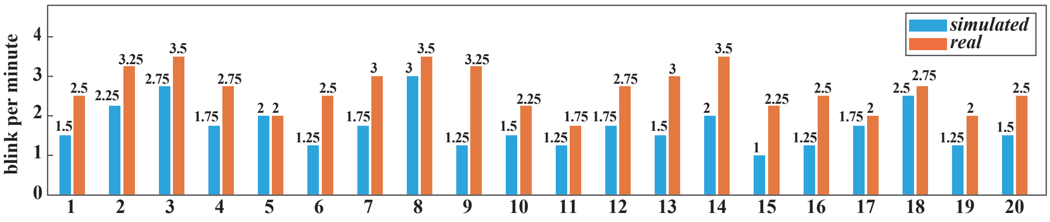


Figure 3.20: Performance of eye blink detection

Performance of yawning, eye blink, and ECD detection under masks and glasses

During driving, the driver may wear a mask for health concerns or sunglasses to protect the eyes. However, the yawning and eye blink features when wearing masks and sunglasses can hardly be detected using cameras. Whereas one of the key advantages of the mmDrive is that the mmWave radar signal can detect the driver's facial features even when the face is occluded. This is because the mmWave radar signal can penetrate through the thin mask and lens. In this evaluation, we investigate the performance of mmDrive to detect yawning and eye blink features while we ask two volunteers to wear disposable medical masks and commonly used glasses(metal frame eyes, resin frame glasses, sunglasses) during driving. The average accuracy of yawning detection acc_y when wearing masks is 96%, which is only around 1% lower than that without masks. Furthermore, the average MAE of eye blink rate are around 3.5 and 2.6 and 2.1 blink per minute, and the average error of ECD estimation are 40.5ms, 38.5ms and 37.5ms, which are all comparable with the results without wearing glasses.

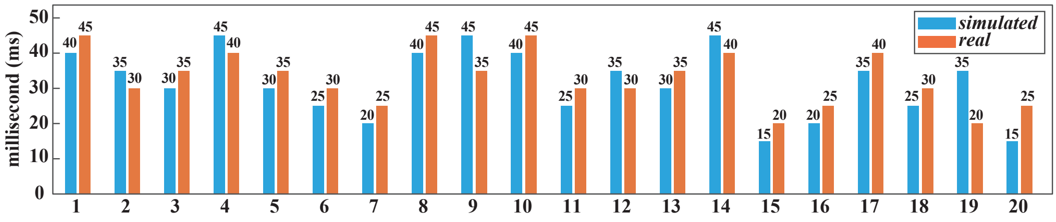


Figure 3.21: Performance of ECD measurement

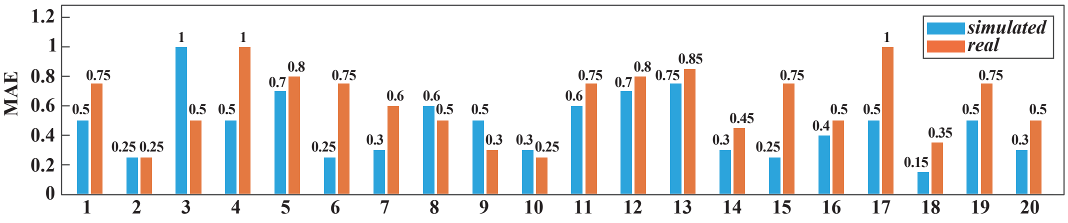


Figure 3.22: Performance of respiration rate estimation

Comparison with existing works

We compare the performance of mmDrive for detecting all target fatigue features with existing works in Table 3.5. For nodding detection, mmDrive can achieve similar accuracy as the camera-based approach of around 97.58% and slightly higher accuracy than the RFID-based method. For yawning detection, we also achieve a comparable accuracy of around 97.4% as existing camera-based methods. In terms of eye blink detection, we first convert the MAE of eye blink rate into the accuracy of successfully detected eye blinks over the total number. The accuracy for eye blink detection using our method is $2\% - 4\%$ smaller than that of existing methods using the camera and sonar. This is mainly because the eye blink movement is too tiny. Albeit our signal processing methods eliminate the effect from other movements, the relative distance between the radar and the driver may still cause some slight blinks to be undetected. For the physiological signals, the MAEs of respiration and heartbeat rate estimation of mmDrive are in a similar range as existing systems using WiFi and mmWave signals.

In a nutshell, the performance of mmDrive to detect fatigue features is comparable

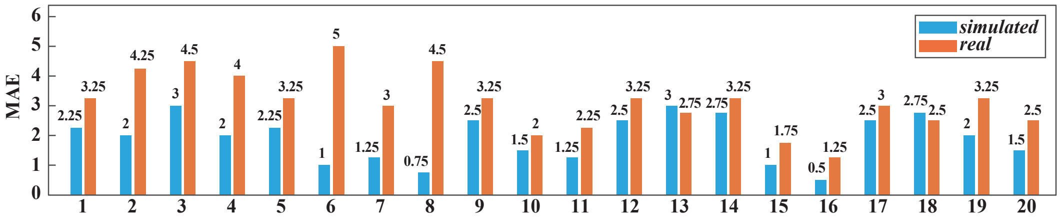


Figure 3.23: Performance of heartbeat rate estimation

with existing works. In addition, mmDrive can detect multiple fatigue features using a single and compact mmWave radar. Whereas most existing systems, e.g., camera-based, RFID-based, or WiFi-based, fail to capture such a whole list of fatigue features from different body regions.

Effect of road conditions

In this evaluation, we investigate the effect of different road conditions on the performance of fatigue feature detection. The detection results for the smooth highway and urban road are given in Table 3.6. The detection performance on the smooth highway is slightly higher than that on the urban road. This is mainly because the driver is more likely to stay in a quasi-static state on the long and smooth highway, while the radar signal suffers from more noise from the relatively rough urban road.

For roads with obvious uneven surfaces, the vehicle would vibrate on a large scale, which incurs the radar and the driver's body to vibrate accordingly. As we have shown in Figure 3.9, the vehicle bump caused by a deceleration zone results in a significant phase change of the signal from the chest bin. Such a large signal change largely overwhelms the tiny facial and physiological features. In our system, we will detect these rough road conditions and skip the fatigue feature detection for the time being.

Table 3.5: Performance comparison with existing works

Feature	Existing works		mmDrive
	Method	Result	Our Result
Nodding	camera [123]	95.7%	97.58%
	RFID [130]	94.3%	
Yawning	camera [66, 34]	95% - 97%	97.4%
Eye blink	camera [61]	97%	93.4%
	sonar [73]	95%	
Respiration	WiFi [19]	$0.3bpm$	$0.54bpm$
	RFID [131]	$0.3\text{-}0.5bpm$	
Heartbeat	WiFi [19]	$4bpm$	$2.52bpm$
	mmWave [92]	$3bpm$	

Performance of fatigue determination algorithm

In this subsection, we compare rule-based and four learning-based fatigue determination algorithms.

We first compare rule-based and four learning-based fatigue determination algorithms for the training set and validation set from the same source. The specific comparison is shown in Table 3.7. It can be seen from the results that the accuracy and F1-score of the rule-based algorithm are inferior to the other four learning-based methods. Among the four learning-based methods, the accuracy and F1-score of the NN model are the highest, while the accuracy and F1-score of the NB method are the lowest, respectively for 98.95% and 98.27%. While the accuracy and F1-score of the rule-based method are 96.5% and 96.49%, respectively, lower than all four learning-based methods.

Table 3.6: Performance of fatigue feature detection under different road conditions

Condition	Nod	Yawn	BlinkR	ECD	RR	HR
Highway	96%	97.5%	$1.5bpm$	$28ms$	$0.4bpm$	$1.65bpm$
Urban road	94.5%	96.2%	$1.9bpm$	$36ms$	$0.45bpm$	$1.88bpm$

Table 3.7: Comparison of fatigue determination algorithm (same source training and validation data-set)

Algorithm	TP	TN	FP	FN	Accuracy	F1-score
Rule-based	1923	1937	63	77	96.5%	96.49%
SVM	1970	1981	19	30	98.78%	98.77%
KNN	1964	1972	28	36	98.4%	98.40%
NB	1954	1977	23	46	98.28%	98.27%
NN	1982	1979	24	18	98.95%	98.95%

Then, we compare rule-based and four learning-based fatigue determination algorithms' ability to address individual differences. The specific comparison is shown in Table 3.8.

Among the four learning-based methods, the accuracy and F1-score of the NN model are the highest, respectively 97.63% and 97.62%, while the accuracy and F1-score of the NB method are the lowest, respectively for 93.75% and 93.72%. Among the four methods we tried, the accuracy rates of the KNN and NB methods are all lower than 95%.

The results show that the accuracy rate and F1-score of the rule-based method proposed in our study are 95.40% and 95.39%, respectively, surpassing the performance of the KNN and NB learning-based methods. Therefore, we posit that the issue of

Table 3.8: Comparison of fatigue determination algorithm ((different source training and validation data-set)

Algorithm	TP	TN	FP	FN	Accuracy	F1-score
Rule-based	1903	1913	87	97	95.40%	95.39%
SVM	1930	1919	81	70	96.23%	96.24%
KNN	1887	1896	104	113	94.58%	94.56%
NB	1865	1885	115	135	93.75%	93.72%
NN	1947	1958	45	53	97.63%	97.62%

individual differences may not be effectively addressed by learning-based methods in our study due to the limited number of samples available for analysis. The rule-based method we proposed is based on a comprehensive literature review, so the number of samples is much larger than the data we collect for machine learning. We also use the pre-collected data of each volunteer while driving as the benchmark data to adjust our algorithm to better deal with the problem of individual differences. In the rule-based method, the fusion of multiple features is not tightly integrated, as it only takes into consideration the occurrence of a single facial feature. However, even if we have a small number of samples, the accuracy rate of the NN model has reached 97.63%, and we believe that the effect will be better when more samples are added.

Based on the above analysis, we believe that the two methods can be combined. We first train a NN model for driver fatigue determination. During the actual driving process, we continuously collect the data of the driver under normal conditions. Then we generate some fatigue data according to the features changes that appear in the driver's fatigue state summarized in the rule-based method. Finally, we continue to fine-tune the model with new data to make the model better adapt to individual differences.

3.5 Related Work

In this section, we introduce the related works for fatigue driving detection, which can be categorized into vision-based, wearable-based, and RF-based approaches.

3.5.1 Vision-based Approaches

Vision-based approaches commonly utilize cameras and infrared sensors to detect various facial fatigue features such as nodding, yawning, eye blink, and gaze direction [98, 20, 2, 65, 127]. Real-time images captured by cameras are processed using machine learning algorithms, such as back propagation neural network (BPNN) and convolutional neural network (CNN), to classify different facial regions and corresponding fatigue features. Infrared (IR) sensors, mounted on the headrest of the driver's seat, are also employed to monitor head position by measuring changes in distance between the specified head region and the IR signal receiver to detect head movements[63]. However, vision-based approaches have fatal limitations including performance degradation in poor lighting conditions such as dimmed light in tunnels or dark environments at night, and blockage of facial regions when drivers wear masks or sunglasses. Additionally, concerns about privacy leakage of facial images may arise among drivers.

3.5.2 Wearable-based Approaches

Wearable sensors have been designed and investigated for fatigue driving detection, with many of them focusing on capturing the driver's physiological signals. For example, wearable systems such as FEELythm developed by Fujitsu [80], are attached to the driver's earlobe to detect the pulse rate during driving. Chest belt devices like BioHarness [121], are used to measure respiration and heartbeat rates. Wearable EEG devices worn on the driver's head can measure EEG signals, which can be analyzed using data mining techniques, for early warning of fatigue driving [143]. Additionally, smart glasses with proximity sensors can be utilized for eye blink detection [14]. While these wearable sensors are effective in capturing physiological responses during driving, they may be inconvenient for drivers to wear for long periods of time. Furthermore, there is a lack of wearable sensors that can simultaneously detect various facial fatigue features, which may weaken the robustness of fatigue driving detection if relying solely on physiological signals.

3.5.3 RF-based Approaches

RF human sensing, which utilizes the signal reflection effect from the human body, has been applied for fatigue driving detection in recent years. Various RF signals, such as WiFi [52, 8], RFID[130, 131], and radar signals [141, 138], have been explored for this purpose.

Among different RF signals, long-wavelength RF signals (i.e., WiFi and RFID) are mainly capable of detecting coarse-grained fatigue features. For example, WiFind employs the channel state information (CSI) of the WiFi signal, which involves the driver's body movement information, to detect the nodding, yawning, and respiration activities [52]. In RFID-based systems, RFID tags are attached to the human body, especially on the chest, to obtain the respiration rate during driving [130]. However, WiFi and RFID signals have limitations such as long wavelength and narrow bandwidth, which restrict their ability to detect fatigue features from multiple body regions simultaneously.

In contrast, short-wavelength and wide-band RF signals (i.e., UWB and mmWave radar) enable the measurement of fine-grained features from different body regions. For instance, in studies utilizing UWB radar, V2iFi employs a compact UWB radar to extract the subtle heartbeat rate variability pattern from the driver [141], and Ubi-Fatigue can detect both eye blinks and heartbeats using a 60GHz radar [138].

However, these UWB radar signals are limited by having only a single Tx-Rx pair, which constrains spatial sensing in the angular dimension, despite their high range resolution for separating signals from different ranges. As a result, signals from the same range but different angles, such as head and hand-reflected signals, are challenging to separate. In studies utilizing mmWave radar, angle-related challenges can be mitigated, but there are still limitations. For example, Liu et al. [75] proposed a method that detects only the heartbeat signal and used machine learning for fatigue detection, which may not capture the full complexity of fatigue-related physiological changes. Dong et al. [30] and Ciattaglia et al. [23] also focused on detecting respiration and heartbeat signals using mmWave radar, which may not be sufficient as the main feature of fatigue was reflected on the driver's face. Shu et al. [97] proposed an eye blink detection method using frequency-modulated continuous wave radar, and Cardillo et al. [12] utilized mmWave radar to detect head and eye blink movements, but their method only detected individual blinks and did not use it for fatigue detection. Similarly, Raja et al. [87] proposed a technique for in-vehicle driver/passenger 3D head tracking using mmWave radar, specifically targeting three different head movements, but the focus was primarily on tracking the position and angle of head movements, and fatigue-related behaviors such as nodding was not taken into account. Sen et al. [95] used range-Doppler information from mmWave radar to detect specific behaviors such as left-look, right-look, talking, and yawning, but their method did not explain how to use these features for driver status determination and only yawning was directly related to fatigue.

In our paper, we employ the mmWave radar, which owns a full advantage of short wavelength, wide band, and multiple Tx-Rx pairs, to renovate fatigue driving detection. Compared with existing RF-based systems, our system achieves the accurate detection of the highest number of fatigue features (four facial and two physiological features) from both head and chest regions.

3.6 Discussion

Despite the promising results demonstrated by our mmDrive system, several important limitations must be acknowledged. The current experimental validation, while sufficient for proof-of-concept demonstration, remains limited in scope with approximately 20 volunteer participants and 200 hours of data collection. This scale falls short of fleet-level validation and does not comprehensively cover the diverse range of real-world driving conditions, including varied traffic densities, weather scenarios (rain, fog, snow), and different vehicle types that would be encountered in large-scale deployment. Additionally, our system exhibits sensitivity to environmental factors such as vehicular vibrations, sudden driving maneuvers, and multipath propagation within the confined cabin space, which can degrade the signal-to-noise ratio for subtle physiological features like eyelid motion detection. The extraction of fine-grained fatigue indicators under dynamic driving conditions remains challenging, particularly when distinguishing genuine physiological signals from motion artifacts and environmental interference. Furthermore, while our spatial resolution allows for driver-passenger discrimination under normal conditions, extreme driving scenarios or unusual seating arrangements could potentially compromise this capability.

In the current version of mmDrive, we achieve effective fatigue driving detection using the mmWave radar. Although we can detect various fatigue features, we can only identify the fatigue state after the presence of fatigue features. It would be desirable if we could utilize the transition and change of the eye blink rate, ECD, and physiological signals to predict whether the driver feels fatigued as early as possible. In our future work, we will investigate the usage of mmWave radar to capture more behavioral characteristics and combine these transitional fatigue features for early prediction of the fatigue state. Apart from fatigue driving, we will also investigate the mmWave radar for detecting other dangerous driving behaviors and integrate them into a mmWave-based safe driving system.

3.7 Conclusion

In this paper, we present our designed mmDrive system for fatigue driving detection. Different from conventional approaches that employ intrusive cameras and bulky wearable sensors, our system only uses a compact mmWave radar to realize non-intrusive yet fine-grained fatigue driving detection by sensing various fatigue features (i.e., nodding, yawning, eye blink, eye closure duration, respiration, and heartbeat) from the driver's head and chest regions. We dedicatedly model the effect of different fatigue features on the mmWave radar signal and develop effective signal processing methods to accurately detect target features from the noisy signals and other interfering activities. mmDrive will determine whether the driver is fatigued or not based on the detected features. Experimental results in both simulated and real driving environments show that mmDrive can achieve high accuracy in fatigue detection.

Chapter 4

LiquidAuth: Reliable and Accurate Liquid Authentication Using GAN-enhanced Acoustic-to-Mass-Spectrum Mapping

Abstract

Counterfeit and adulterated liquids present significant health risks and economic losses, underscoring the need for effective authentication methods. While acoustic signal-based detection offers a promising non-invasive approach that works without opening containers, its accuracy suffers from variations in container properties and environmental conditions. Furthermore, acoustic features alone lack molecular-level detail needed for definitive identification. We address these challenges with LiquidAuth, a system that maps acoustic signals to mass spectra, providing molecular-level insights for more accurate authentication. LiquidAuth employs a cross-shaped microphone array to mitigate positional variation and introduces an adaptive container compensation algorithm to account for different container characteristics. By integrating Conditional GANs (cGANs), our system effectively maps acoustic signals to mass spectra, enabling reliable molecular-level classification. Experimental evaluations show LiquidAuth achieves an average F1-score of 97.89%, with accuracy between 95.35% and 98.25% across various container materials and storage conditions, demonstrating robust liquid authentication capabilities.

4.1 Introduction

Counterfeit and adulterated liquids represent a significant threat to public health and economic stability, causing numerous poisoning incidents and substantial financial losses annually [82]. The detection and verification of liquid authenticity is therefore critical; however, conventional methodologies including chemical analysis, chromatographic techniques, Quasi-Static Electrical Tomography (QET), and surface tension measurement necessitate container opening and specialized equipment, thereby compromising product integrity and increasing operational costs [109, 135, 39, 41]. Despite their efficacy in contaminant detection, these intrusive approaches fail to provide scalable, non-destructive solutions. Recent advancements in non-invasive technologies, such as RFID-based systems [39], ultra-wideband (UWB) approaches [48], and millimeter-wave (mmWave) radar systems like FG-LiquID [70], offer promising alternatives by preserving container integrity. Nevertheless, these systems exhibit sensitivity to environmental interference and require specialized hardware, limiting their scalability.

Acoustic-based methodologies [133] present a cost-effective, non-invasive alternative by exploiting the distinctive absorption and transmission characteristics of sound waves in various liquids. These systems utilize strategically positioned speakers and microphones around the container to measure alterations in sound wave propagation, subsequently analyzing the Acoustic Absorption and Transmission Curve (AATC) for liquid identification and counterfeit detection.

Despite their potential, existing acoustic methodologies exhibit significant limitations when confronted with variations in container materials, geometries, and environmental conditions, while simultaneously lacking molecular-level precision essential for definitive liquid identification—thus compromising both accuracy and reliability. Environmental fluctuations, container material heterogeneity, and positional variations between acoustic transducers introduce substantial measurement uncertainties [133]. Furthermore, while sound wave absorption patterns reflect certain liquid characteristics, they lack the interpretability and precision offered by chemical methodologies, potentially undermining detection reliability.

In this paper, we propose **LiquidAuth**, an innovative non-invasive liquid authentica-

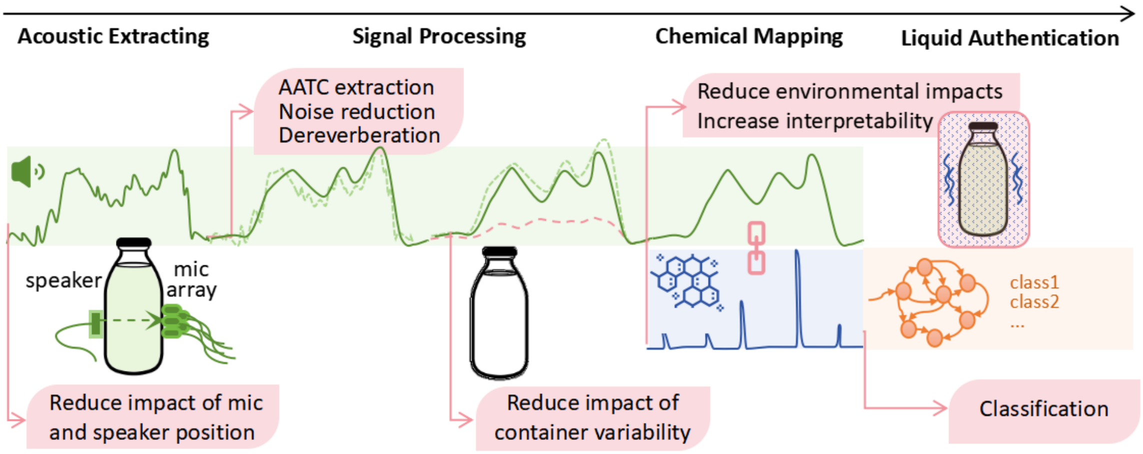


Figure 4.1: System architecture of LiquidAuth, illustrating the novel acoustic-to-mass-spectrum mapping approach for molecular-level liquid identification.

tion system that introduces acoustic-to-mass-spectrum mapping to extract molecular-level details from liquids. Unlike traditional acoustic-based detection methodologies, which primarily rely on acoustic features, our approach captures microscopic molecular characteristics by inferring mass spectra from acoustic signals, thereby significantly enhancing the accuracy of fine-grained liquid identification. However, accurately transforming acoustic signals into mass spectra presents substantial challenges due to environmental variations, container effects, and the inherent complexity of the signal relationships. To address these challenges, LiquidAuth employs several advanced techniques, including a Generative Adversarial Network (GAN) for mapping acoustic features to precise mass spectra. The system architecture of LiquidAuth is illustrated in Fig. 4.1.

We implement several innovative approaches to tackle the technical challenges inherent in acoustic liquid authentication. We deploy a cross-shaped array of five microphones surrounding the container, complemented by a speaker to mitigate positional variation and multipath effects. This configuration captures signals from multiple angles and substantially reduces the impact of positional shifts during measurement. We further enhance signal quality through advanced signal processing techniques, including adaptive noise reduction algorithms and dereverberation methods, which significantly improve signal clarity and fidelity for subsequent analysis. Second, recognizing that variations in container material and geometry can significantly interfere with acoustic transmission, we develop a container compensation model that effectively normalizes signals based on container characteristics, thus ensuring consistent performance across diverse container types. Finally, we implement Conditional Generative Adversarial Networks (cGANs) to accurately map acoustic signals to molecular-level mass spectra, subsequently applying a Fully Connected Neural Network (FCNN) to classify the generated spectra. This approach enables precise molecular-level detection and provides reliable liquid authentication even under varied storage conditions.

In summary, we make the following key contributions:

- We propose **LiquidAuth**, the first system to map acoustic signals to mass spectra for liquid authentication, integrating advanced signal processing techniques and cGANs to achieve high accuracy and interpretability across various storage conditions.
- We design a cross-shaped microphone array and a container compensation model to mitigate positional variation as well as variations in container material and shape, enhancing the accuracy of acoustic signal analysis and system generalizability.
- Extensive experiments, including a case study on alcoholic beverages, demonstrate that LiquidAuth accurately distinguishes between different types and brands of liquids, even capturing subtle differences such as wine from different production years.

The remainder of this paper is organized as follows: Section II provides the theoretical background on the relationship between acoustic signals and mass spectra for liquid authentication. Section III gives an overview of the LiquidAuth system architecture.

Section IV details the technical methodologies, including signal processing, hardware design, container compensation, and the cGAN-based mapping from AATC to mass spectra. Section V describes the experimental setup and deployment strategies. Section VI presents the evaluation results, focusing on system performance across different liquid types and conditions. Section VII reviews related work. Section VIII discusses key challenges, limitations, and future improvements. Finally, Section IX concludes the paper with key findings and future research directions.

4.2 Preliminaries

In this section, we explore the principles behind the acoustic signal and its relationship with mass spectra, which are the primary features utilized in our system.

4.2.1 Relationship Between Acoustic Properties and Molecular Composition

When acoustic waves travel through a liquid, they interact with the liquid's molecular structure, affecting the wave's absorption and transmission characteristics. The acoustic impedance Z of a liquid is influenced by its density ρ and the speed of sound c in the medium [26]:

$$Z = \rho \cdot c. \tag{4.1}$$

These parameters are directly related to the liquid's molecular composition. Variations in molecular structure lead to differences in viscosity and elasticity, which in turn affect the acoustic absorption coefficient $\alpha(f)$ at frequency f:

$$\alpha(f) = \frac{2\pi^2 f^2 \eta}{\rho c^3}, \tag{4.2}$$

where η is the dynamic viscosity [84]. By analyzing the frequency-dependent acoustic signal, we can obtain information related to the molecular properties of the liquid.

4.2.2 Acoustic Absorption and Transmission Curve (AATC)

Based on a recent work [133], we utilize the Acoustic Absorption and Transmission Curve (AATC), a feature extracted from the acoustic signal as it travels through the liquid, to capture key acoustic characteristics for liquid authentication.

The emitted acoustic signal $s(t)$ consists of multiple sine waves with amplitudes A_i at different frequencies f_i ranging from 18 kHz to 20 kHz:

$$s(t) = \sum_{i=1}^{n} A_i \sin(2\pi f_i t), \tag{4.3}$$

where n is the number of discrete frequencies. These high frequencies are chosen to minimize background noise and human voice interference [64].

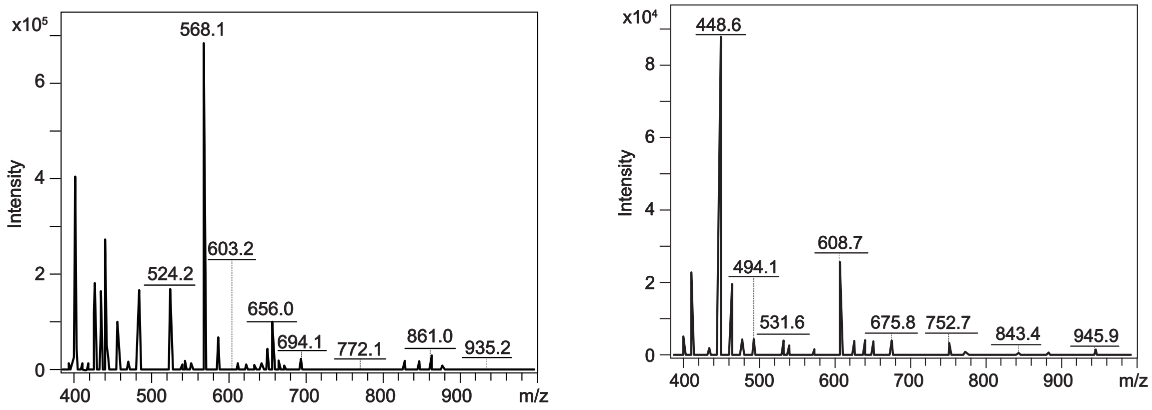
After transmitting through the liquid, the received signal is captured by the microphones. We perform signal processing, including noise reduction and dereverberation, to extract the amplitudes $R(f_i)$ at frequencies f_i. The AATC at each frequency f_i is then calculated as:

$$\text{AATC}(f_i) = \frac{R(f_i)}{S(f_i)}, \tag{4.4}$$

where $S(f_i) = A_i$ are the amplitudes of the emitted signal at frequencies f_i. The AATC reflects the frequency-dependent absorption characteristics of the liquid, which are influenced by its molecular composition.

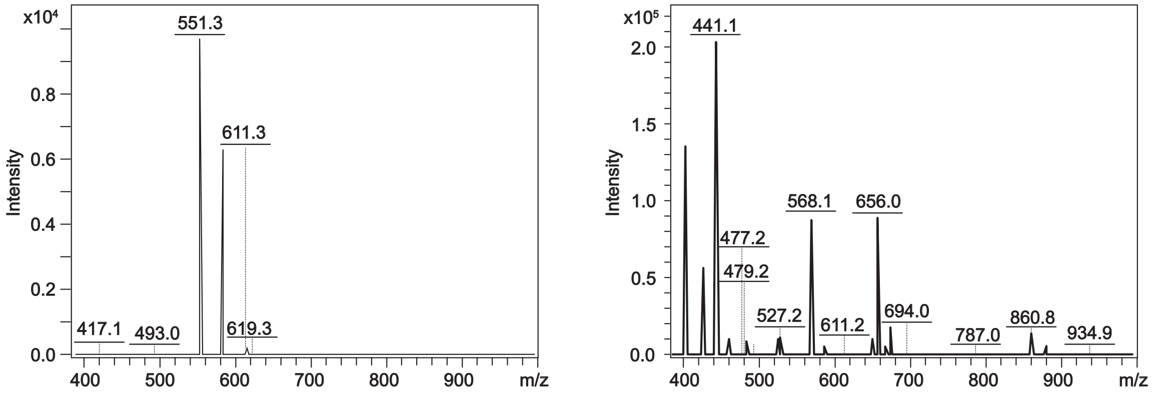
4.2.3 Mass Spectrometry and Molecular Fingerprinting

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry provides detailed information about the molecular composition of a sample by measuring the mass-to-charge ratio (m/z) of ionized particles [72]. Each liquid has a unique mass spectrum that serves as a molecular fingerprint, as shown in Fig. 4.2.



(a) wine

(b) beer



(c) whisky

(d) vodka

Figure 4.2: MALDI-TOF mass spectra of alcoholic beverages.

4.2.4 Bridging Acoustic Features and Mass Spectra

As mentioned previously, the connection between acoustic absorption characteristics and mass spectra lies in their sensitivity to a liquid's molecular composition. Both of them respond to variations in molecular structure, influencing properties such as density, viscosity, and elasticity, which impact acoustic absorption and ionization patterns in mass spectrometry. While AATC provides macroscopic acoustic information influenced by molecular properties, mass spectrometry offers microscopic molecular composition details. Our key innovation is leveraging a data-driven approach using

Conditional GANs (cGANs) to map the AATC to the corresponding mass spectra. This mapping is feasible because variations in acoustic behavior, influenced by molecular composition, correspond to molecular signatures detected through mass spectrometry. It allows us to infer molecular-level information from acoustic measurements, enhancing the interpretability and accuracy of liquid authentication.

4.3 System Design

In this section, we present the architecture of LiquidAuth, highlighting the key components and workflow of the system, as illustrated in Fig. 4.3.

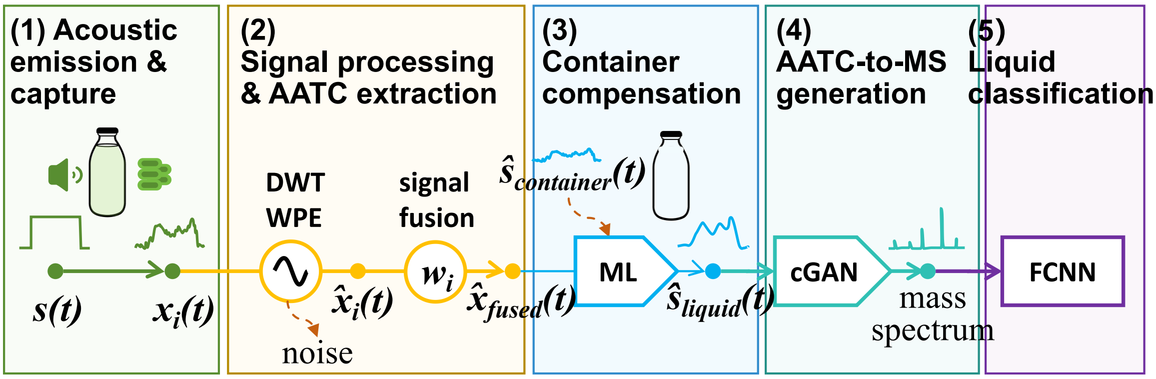


Figure 4.3: System architecture of LiquidAuth.

The LiquidAuth system consists of the following main components:

1. **Acoustic Signal Transmission and Reception**: A speaker emits designed acoustic signals through the liquid-filled container. To capture the transmitted signals, we use a cross-shaped array of five microphones (M1 to M5), arranged around the container to mitigate multipath effects and positional variability, as shown in Fig. 4.4. Soundproofing materials surround the speaker and microphones to minimize external noise interference.

2. **Signal Processing and AATC Extraction**: The received signals undergo

advanced signal processing techniques, including noise reduction and dereverberation, to enhance signal clarity. We then extract the Acoustic Absorption and Transmission Curve (AATC) from the processed signals, which reflects the frequency-dependent absorption characteristics of the liquid.

3. **Container Compensation Model**: To address the variability introduced by different container materials and shapes, we implement a container compensation model. By collecting data from various containers and applying machine learning algorithms, the model adjusts the AATC to minimize the container's impact on the acoustic measurements.

4. **Acoustic-to-Mass-Spectrum Mapping Using cGAN**: The compensated AATC features are input into a trained Conditional GAN, which maps them to corresponding mass spectra. This mapping allows us to infer molecular-level information from the acoustic data, enhancing the interpretability and robustness of the system.

5. **Liquid Authentication via Classification**: The generated mass spectra are analyzed using a classification model, such as a Fully Connected Neural Network (FCNN), to determine the authenticity and type of the liquid.

4.4 Methodology

In this section, we detail the design of the LiquidAuth system and the key challenges addressed. The main challenges include mitigating multipath effects from microphone and speaker positional variations, eliminating container influence on AATC extraction and overcoming different storage environment impacts on liquid quality. We address these through innovative solutions: using a microphone array for signal extraction and fusion, advanced signal processing techniques for AATC extraction, a container compensation model, and a conditional Generative Adversarial Network (cGAN) to map AATC features to mass spectra. Finally, we employ a Fully Connected Neural Network (FCNN) for liquid authentication based on the generated mass spectra.

4.4.1 Signal Extraction and Fusion Using a Microphone Array

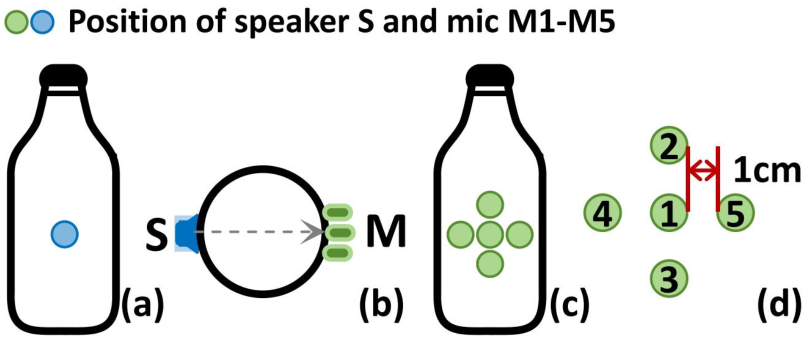


Figure 4.4: Position of speaker and microphones: (a) speaker view; (b) top view; (c) microphone view; (d) 1 cm spacing between microphones.

As shown in Fig. 4.4, we designed a cross-shaped array of five microphones (M1 to M5) strategically arranged around the container:

- Microphone M1 directly faces the speaker S, aligned with the maximum container diameter.
- Microphones M2 and M3 are placed directly above and below M1, respectively.
- Microphones M4 and M5 are placed to the left and right of M1, respectively.
- The interval between each microphone is 1 cm.

Inside a liquid container, emitted sound waves undergo reflections from both container walls and the liquid surface, producing multipath components that arrive at microphones with varying delays and amplitudes. These multipath copies distort the waveform and degrade the accuracy of AATC estimation. To address this, we integrate multipath suppression with microphone array fusion as follows.

We first locate the direct arrival path by computing the cross-correlation between the transmitted signal $s(t)$ and each received signal $x_i(t)$, where the first correlation peak marks the arrival time $\tau_{0,i}$. Around this reference, three windows are defined: the direct window capturing the main path,

$$W_{\text{dir}} = [\tau_{0,i} - \Delta,\ \tau_{0,i} + \Delta],$$

the echo window containing delayed reflections,

$$W_{\text{echo}} = [\tau_{0,i} + \Delta,\ \tau_{\max}],$$

and a pre-emission window serving as the baseline noise segment,

$$W_{\text{pre}} = \text{a short pre-emission segment.}$$

Using these windows, we compute the energies for each microphone i:

$$E^{(i)}_{\text{direct}} = \sum_{t \in W_{\text{dir}}} x_i^2(t), \quad E^{(i)}_{\text{noise}} = \sum_{t \in W_{\text{pre}}} x_i^2(t),$$

$$E^{(i)}_{\text{echo}} = \max\left(\sum_{t \in W_{\text{echo}}} x_i^2(t) - E^{(i)}_{\text{noise}},\ 0 \right).$$

From these, an effective SNR is defined to characterize the reliability of each channel:

$$\text{SNR}_i = \frac{E^{(i)}_{\text{direct}}}{E^{(i)}_{\text{echo}} + E^{(i)}_{\text{noise}} + \varepsilon},$$

where ε is a small constant preventing division by zero. Intuitively, multipath increases $E^{(i)}_{\text{echo}}$, thereby reducing SNR_i.

With these SNR values, we assign adaptive fusion weights to the microphones:

$$w_i = \frac{\text{SNR}_i}{\sum_j \text{SNR}_j}, \qquad x_{\text{fused}}(t) = \sum_{i=1}^{5} w_i\, x_i(t).$$

In this way, microphones with cleaner direct paths are emphasized, while those dominated by echoes or noise are downweighted. The fused signal is then transformed into the frequency domain using FFT, and normalized by the transmitted reference signal to obtain AATC(f). The resulting spectrum reflects the intrinsic absorption characteristics of the liquid with multipath distortions effectively suppressed.

By dynamically recomputing w_i for each emission, the system adapts to varying acoustic conditions, substantially enhancing the robustness and precision of AATC estimation.

4.4.2 Signal Processing and AATC Extraction

To ensure the fused signal $x_{\text{fused}}(t)$ accurately represents the liquid's acoustic characteristics, we apply noise reduction and dereverberation processing.

For noise reduction, we employ a Discrete Wavelet Transform (DWT)-based algorithm that decomposes the signal into multiple frequency bands and applies thresholding to wavelet coefficients, effectively preserving signal features while suppressing environmental noise and electromagnetic interference.

We address reverberation using the Weighted Prediction Error (WPE) algorithm, which models and subtracts late reverberation components to enhance the direct-path signal, improving temporal resolution and frequency analysis accuracy.

Following preprocessing, we extract the AATC through:

1. Fast Fourier Transform of $x_{\text{fused}}(t)$ to obtain $X_{\text{fused}}(f)$
2. Amplitude extraction at frequencies 18-20 kHz: $R(f_i) = |X_{\text{fused}}(f_i)|$
3. Normalization by the emitted signal amplitudes: $\text{AATC}(f_i) = \frac{R(f_i)}{S(f_i)}$

The resulting AATC captures frequency-dependent absorption characteristics influenced by the liquid's molecular properties, providing the foundation for subsequent composition analysis.

4.4.3 Container Compensation Model

Container material properties (density, elasticity), shape, and thickness can significantly alter acoustic signal transmission and attenuation, potentially distorting the liquid's true acoustic signature and leading to inaccurate AATC measurements. To address this issue, we develop a container compensation model that explicitly quantifies and corrects for container influence.

We first collect comprehensive AATC data from liquids across a wide variety of container types, including materials (glass, plastic, metal) and shapes (cylindrical, rectangular, irregular). This dataset captures diverse transmission effects and provides the basis for training a compensation model. Formally, we define a container transfer function $H_c(f)$ as

$$H_c(\mathrm{AATC}_{\mathrm{container+liquid}}(f)) = \mathrm{AATC}_{\mathrm{liquid}}(f), \tag{4.5}$$

where $\mathrm{AATC}_{\mathrm{container+liquid}}(f)$ is the measured spectrum and $\mathrm{AATC}_{\mathrm{liquid}}(f)$ is the intrinsic liquid response from reference measurements.

To approximate $H_c(\cdot)$, we employ a neural network with container characteristics and measured AATC as input, nonlinear hidden layers, and an output layer predicting the corrected container-free AATC. Training is guided not only by mean squared error between predicted and reference curves, but also by a contrastive learning strategy. In this strategy, positive pairs consist of the same liquid measured in different containers, for which the corrected outputs should be close:

$$D(E(x_a)) \approx D(E(x_b)) \quad \text{if same liquid, different containers.}$$

Negative pairs are liquids of different types, which must remain discriminative after compensation:

$$D(E(x_a)) \neq D(E(x^-)) \quad \text{if different liquids.}$$

This learning objective enforces both intra-liquid consistency and inter-liquid separability, ensuring that compensation does not collapse class distinctions. At inference time, the trained model takes a new container-influenced AATC x and outputs its container-free version $\hat{x}$, which can then be used for downstream tasks such as mass spectrum mapping or liquid authentication.

By combining regression with contrastive constraints, the model effectively removes container-induced distortions and generalizes to new container types. This enhances LiquidAuth's robustness in real-world scenarios, allowing accurate and reliable authentication across diverse packaging without significant accuracy loss.

4.4.4 Acoustic-to-Mass-Spectrum Mapping Using cGAN

Establishing a direct mapping between acoustic attenuation characteristics and the mass spectrum is complex due to the nonlinear and high-dimensional relationships involved.

To model this complex relationship, we employ a Conditional Generative Adversarial Network (cGAN), which extends the traditional GAN by conditioning the generation process on additional information (in this case, the AATC). The cGAN consists of two main components:

- *Generator G*: A neural network that takes the AATC x as input and generates a synthetic mass spectrum $\hat{y} = G(x)$.
- *Discriminator D*: A neural network that evaluates the authenticity of the mass spectrum given the AATC, outputting a probability that a pair (x, y) is real or fake.

We prepare a dataset of paired AATC and corresponding mass spectra obtained from controlled experiments. The mass spectra are preprocessed to align peaks and normalize intensities, ensuring consistency across samples. The generator and discriminator are trained simultaneously in an adversarial manner. The generator G aims to produce mass spectra that are indistinguishable from real ones, given the input AATC. The discriminator D attempts to correctly classify real and generated mass spectra, conditioned on the AATC. As a result, the generator becomes capable of accurately inferring mass spectra from the AATC.

The loss functions used are:

- Adversarial loss for D:

$$\mathcal{L}_D = -\mathbb{E}_{x,y}[\log D(x, y)] - \mathbb{E}_x[\log(1 - D(x, G(x)))]. \tag{4.6}$$

- Adversarial loss for G:

$$\mathcal{L}_{\text{GAN}} = -\mathbb{E}_x[\log D(x, G(x))]. \tag{4.7}$$

- Reconstruction loss:

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{x,y}[\|y - G(x)\|^2]. \tag{4.8}$$

- Total loss for G:

$$\mathcal{L}_G = \mathcal{L}_{\text{GAN}} + \lambda\mathcal{L}_{\text{MSE}}, \tag{4.9}$$

  where λ is a hyperparameter that balances the adversarial and reconstruction losses.

We optimize G and D alternately using the Adam optimizer with a learning rate of 1×10^{-4}. The networks are trained for 200 epochs with a batch size of 64, minimizing $\mathcal{L}_G$ and $\mathcal{L}_D$ respectively.

The cGAN framework allows us to capture the complex, nonlinear mapping from AATC to mass spectra without explicitly modeling the underlying physics.

4.4.5 Liquid Authentication via FCNN

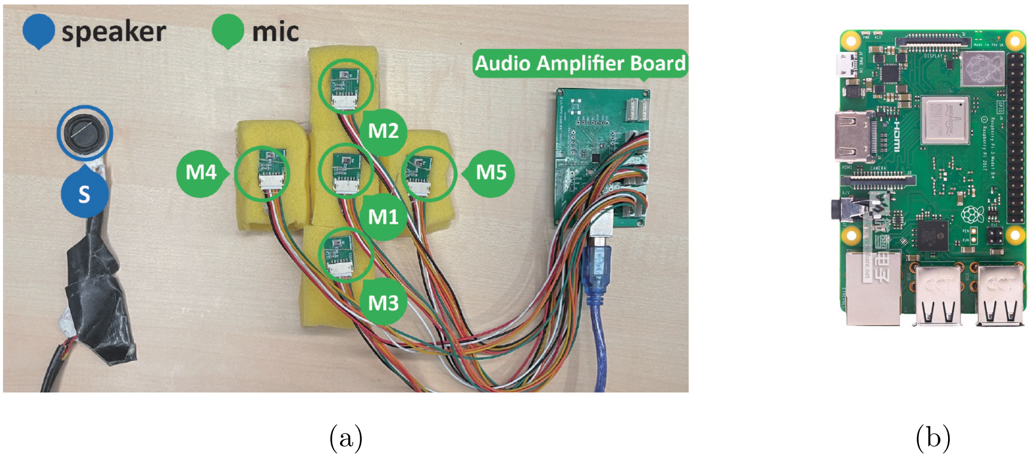
The generated mass spectra contain rich information about the molecular composition of the liquid. To classify these spectra into categories (authentic vs. counterfeit, specific brands, or production years), we employ a FCNN. The FCNN architecture consists of:

- *Input Layer*: Receives the mass spectrum vector, with each element representing the intensity at a specific mass-to-charge ratio (m/z value)
- *Hidden Layers*: Three layers (512, 256, 128 neurons) with ReLU activation functions to introduce nonlinearity
- *Output Layer*: Softmax activation for multi-class classification, outputting a probability distribution over the classes

The network is trained using cross-entropy loss function, which measures the discrepancy between predicted probabilities and true labels. Parameters are updated using the Adam optimizer (learning rate: 10^{-4}), with batch normalization applied after each hidden layer to stabilize and accelerate training. This final classification component leverages the detailed molecular information from the mass spectra to provide accurate liquid authentication across various categories, effectively addressing challenges posed by adulteration and counterfeiting.

4.5 Implementation and Evaluation Metrics

In this section, we describe the hardware and software configurations of LiquidAuth, as well as the experimental settings and evaluation metrics used to assess its performance.



(a) (b)

Figure 4.5: Hardware components used in experiment: (a) Speaker, Microphone Array, Amplifier Board; (b) Raspberry Pi 3B.

4.5.1 Hardware Configuration

Our hardware setup includes a speaker, an amplifier board, a microphone array, and a Raspberry Pi 3B, as shown in Fig. 4.5 and Fig. 4.6. These components are essential for generating and capturing acoustic signals and performing mass spectrometry analysis.

The speaker emits the designed acoustic signals through the liquid container, and the microphone array captures the transmitted signals. The Raspberry Pi 3B handles data acquisition and preliminary processing. The mass spectrometer provides the ground truth mass spectra for training and evaluation. By integrating these components, we establish a stable platform for data collection and analysis.

4.5.2 Software Implementation

Acoustic signals were generated and processed using Python. The signals consist of 21 sine waves ranging from 18 to 20 kHz, saved as WAV files. The cGAN and FCNN models were implemented and trained using PyTorch. Mass spectra were

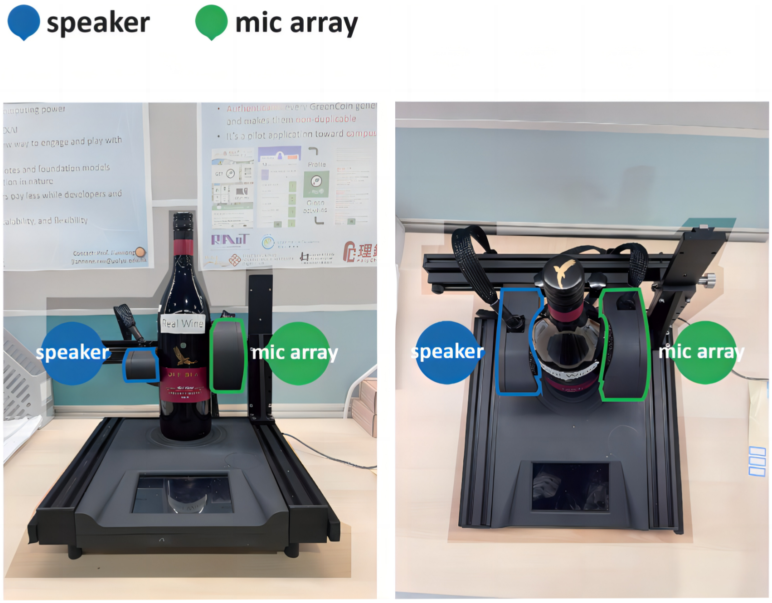


Figure 4.6: LiquidAuth device for data collection.

acquired using the UltrafleXtreme MALDI-TOF mass spectrometer and processed with flexAnalysis software to extract significant peaks and generate the final mass spectra data.

4.5.3 Experimental Settings

To validate LiquidAuth, we tested six alcoholic beverage types: wine (W), beer (B), Chinese white spirit (S), brandy (Br), whisky (Wh), and vodka (V), selecting five popular brands per type (A-E). We created adulterated samples by adding industrial alcohol (30% concentration) at 50%, 25%, and 5% levels, and by diluting with cheaper brands.

To assess robustness, we simulated improper storage conditions including extreme temperatures (8°C, 40°C, 60°C), humidity levels (45%RH, 95%RH), and mechanical vibrations mimicking transport conditions. Samples were labeled as AT_Brand_N, indicating alcohol type, brand, and sample number.

Performance was evaluated using standard classification metrics: Accuracy, Precision, Recall, and F1-score [102], across adulteration detection, brand/year verification, and container compensation effectiveness under both standard and improper storage conditions.

4.5.4 Data Collection

Data collection was conducted under controlled laboratory conditions and simulated environments to mimic real-world storage conditions. For each sample, we collected AATC data and corresponding mass spectra.

AATC Data: Multiple sets of AATC data were collected from each sample over several days to account for variability. Under standard conditions, we collected data over 10 days. For samples under simulated improper storage conditions, we collected additional data to evaluate the system's robustness.

Mass Spectra Data: Samples from the alcoholic beverages were carefully extracted to prevent contamination. Mass spectra were acquired using the mass spectrometer in positive ion mode. Peaks with a signal-to-noise ratio above 3 were extracted to generate the final mass spectra data. These data were used to train and evaluate our cGAN and FCNN models.

4.6 Evaluation Results

In this section, we comprehensively evaluate the performance of LiquidAuth across various tasks, including detecting industrial alcohol adulteration, adulteration with cheaper brands, production year verification, brand identification, and assessing the impact of the container compensation model. We also evaluate the system's robustness under different storage conditions and microphone/speaker arrangements, using metrics such as accuracy, precision, recall, and F1-score to provide a detailed analysis.

4.6.1 Performance of Fraud Detection

Table 4.1 presents LiquidAuth's performance in detecting industrial alcohol adulteration. The system achieves high accuracy across all beverage types, with wine (W) reaching 99.25% under normal conditions and 98.25% under improper storage conditions (*Imp.*), demonstrating resilience to environmental variations. The high precision and recall values indicate effective minimization of both false positives and false negatives.

Table 4.2 shows the system's performance in detecting adulteration with cheaper brands. LiquidAuth maintains high accuracy across all beverage types, with wine achieving an F1-score of 99.50% under normal conditions and 99.25% under improper conditions. This demonstrates the system's ability to detect subtle composition changes from cheaper brand mixing, crucial for maintaining product integrity.

4.6.2 Performance of Brand Detection

Fig. 4.7 summarizes the accuracy of brand detection under normal and improper storage conditions for various types of beverages. LiquidAuth achieves high accuracy across all beverage types, demonstrating its effectiveness in distinguishing different brands based on their unique molecular compositions. For instance, wine (W) achieves an accuracy of 99.25% under normal conditions and 98.25% under improper conditions. This high level of accuracy indicates that the system can reliably identify the correct brand, which is essential for verifying product authenticity and combating counterfeit products. The slight decrease in accuracy under improper storage conditions suggests that while environmental factors can impact performance, the system remains robust and effective.

Table 4.1: Industrial Alcohol Adulteration Detection Performance

Type	**Acc. (%)**	**Prec. (%)**	**Rec. (%)**	**F1 (%)**
W	99.25	99.50	99.00	99.25
B	97.75	99.00	96.59	97.78
S	97.50	98.50	96.57	97.52
Br	96.25	97.00	95.57	96.28
Wh	96.75	97.50	96.06	96.77
V	98.50	99.00	98.02	98.51
W (*Imp.*)	98.25	98.00	98.49	98.25
B (*Imp.*)	97.25	97.00	97.49	97.24
S (*Imp.*)	96.75	98.00	95.61	96.79
Br (*Imp.*)	96.50	98.00	95.15	96.55
Wh (*Imp.*)	96.25	97.00	95.57	96.28
V (*Imp.*)	98.00	99.50	96.60	98.03

Note: *W* = Wine, *B* = Beer, *S* = Chinese white spirit, *Br* = Brandy, *Wh* = Whisky, *V* = Vodka. (*Imp.*) indicates improper storage conditions.

Table 4.2: Cheaper Brand Adulteration Detection Performance

Type	**Acc. (%)**	**Prec. (%)**	**Rec. (%)**	**F1 (%)**
W	99.50	99.50	99.50	99.50
B	97.25	97.00	97.49	97.24
S	98.25	99.50	97.07	98.27
Br	98.00	98.00	98.00	98.00
Wh	99.00	99.00	99.00	99.00
V	98.50	99.00	98.02	98.51
W (Imp.)	99.25	99.50	99.00	99.25
B (Imp.)	97.00	97.00	97.00	97.00
S (Imp.)	98.75	98.00	99.49	98.74
Br (Imp.)	97.75	99.50	96.14	97.79
Wh (Imp.)	98.75	98.00	99.49	98.74
V (Imp.)	98.00	99.50	96.60	98.03

Note: Abbreviations are the same as in Table 4.1.

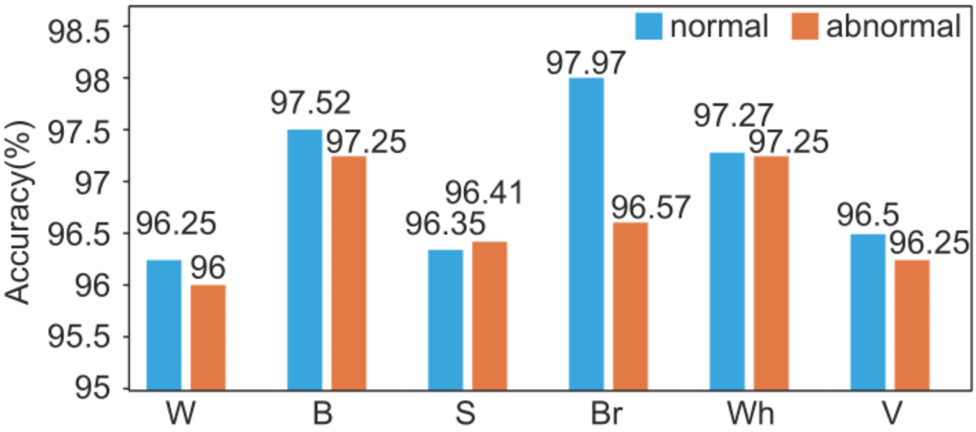
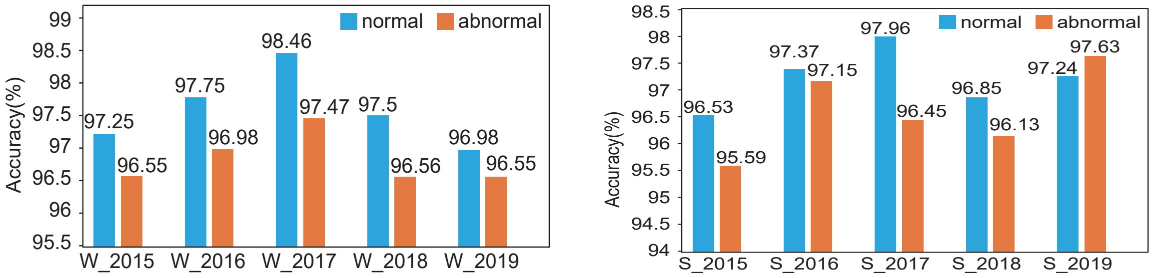


Figure 4.7: Accuracy of brand detection under normal and improper storage conditions for various types of beverages. Abbreviations are the same as in Table 4.1.



(a) Wine Year Verification Accuracy

(b) Chinese White Spirit Year Verification Accuracy

Figure 4.8: Accuracy of year verification under normal and improper storage conditions for wine and Chinese white spirit.

4.6.3 Performance of Year Verification

Fig. 4.8 shows the accuracy of production year verification for wine and Chinese white spirit across different vintages and storage conditions. The system maintains high accuracy, with values remaining above 96% for all tested years. Specifically, for wine, the highest accuracy of 98.46% is achieved for the 2017 vintage. This demonstrates the system's capability to detect subtle differences in the chemical composition of beverages that result from aging processes and production variations over different years. Accurate year verification is particularly important for products like wine and spirits, where the vintage can significantly influence quality and market value. The system's strong performance under both normal and improper storage conditions highlights its reliability for practical applications.

4.6.4 Impact of Container Compensation

We evaluated the system without the container compensation model to assess its impact. As shown in Fig. 4.9, accuracy decreases significantly without compensation, particularly under improper storage conditions. For industrial alcohol adulteration detection, accuracy drops from 95.35% to 93.68%. This decline highlights the impor-

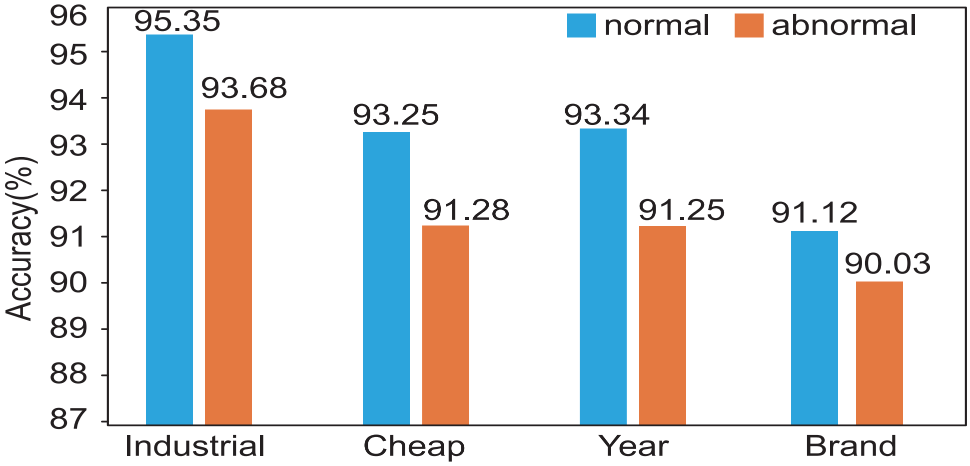


Figure 4.9: Accuracy without container compensation for adulteration under normal and improper storage conditions.

tance of accounting for container-induced distortions that can mask liquid acoustic signatures. The compensation model effectively mitigates these effects, enhancing reliability across various conditions and container types.

4.6.5 Effect of Microphone and Speaker Arrangement

We compared single microphone versus five-microphone array configurations to evaluate hardware impact on detection accuracy. Fig. 4.10 shows the five-microphone array achieves significantly higher accuracy across all tasks, with industrial alcohol adulteration detection reaching 96.67% versus 85.75% with a single microphone under normal conditions. This improvement demonstrates how multiple spatially distributed microphones enhance system robustness by averaging out noise and providing better coverage of the acoustic field around the container.

Table 4.3: Impact of GAN on Detection Accuracy

Task	**With GAN (%)**	**Only AATC (%)**
Year (W, Normal)	97.59	92.75
Year (W, Imp.)	96.82	89.75
Brand (Normal)	96.98	93.60
Brand (Imp.)	96.62	88.20
Adulteration (Normal)	98.04	94.30
Adulteration (Imp.)	97.71	91.50

4.6.6 Ablation Study on GAN's Impact

An ablation study was performed to compare LiquidAuth's performance with and without GAN integration. We focused on three main tasks: year classification, brand classification, and adulteration detection, evaluated under normal and improper storage conditions. Table 4.3 shows that the system consistently performs better with GAN, particularly in fine-grained classification tasks like year verification and under improper storage conditions. For instance, wine year verification accuracy increases from 89.75% to 96.82% with GAN under improper conditions, and adulteration detection improves from 91.50% to 97.71% under improper conditions, demonstrating the GAN's ability to manage environmental variations and subtle differences in liquid composition.

4.6.7 Comparison with Existing Methods

Compared to existing liquid identification methods (92%-97% accuracy), LiquidAuth demonstrates superior performance. FG-LiquID [70] achieves 97% accuracy using 60GHz millimeter-wave radar but may experience performance degradation with container variations. Similarly, MmLiquid [116] reports 97.6% accuracy but requires specialized hardware and complex setups. RFID-based methods like RF-EATS [39] and UWB-based LiquID [48] offer non-invasive detection but show limited accuracy in fine-grained authentication. In contrast, LiquidAuth consistently achieves 98.25% accuracy for wine adulteration detection, even under improper storage conditions, attributable to its advanced container compensation model and cGAN integration, as summarized in Table 4.4.

Table 4.4: Comparison of Liquid Identification Methods

Method	**Accuracy**	**Characteristics**
FG-LiquID [70]	97%	Accuracy may drop with distance and container material variations
MmLiquid [116]	97.6%	Requires specialized hardware and complex setup
RF-EATS [39]	90%	Low accuracy in fine-grained liquid adulteration scenarios
LiquID [48]	94%	Sensitive to container interference and requires complex setups
LiquidAuth (Proposed)	98.25%	Robust in diverse environmental and storage conditions

4.7 Related Work

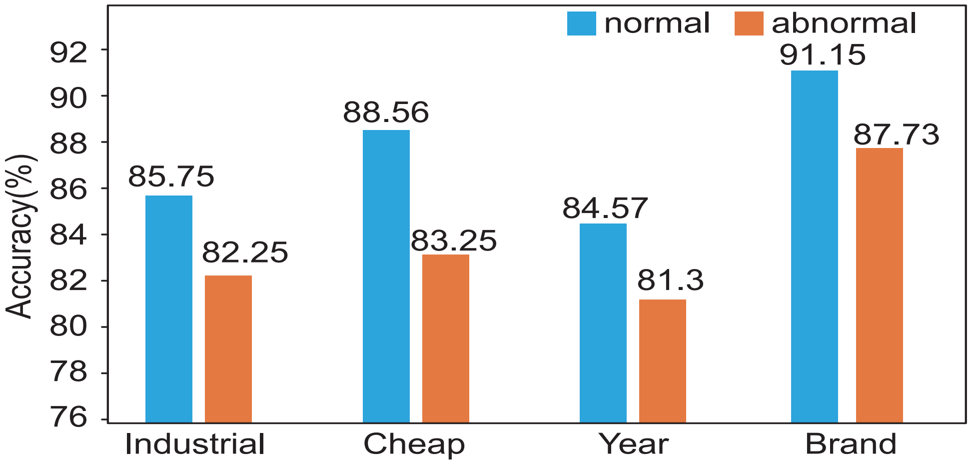
Recent research has explored various techniques for liquid quality detection, including chemical, chromatographic, QET, surface tension-based, RF signal-based, and acoustic methods. Chemical and chromatographic techniques like GC-MS and HPLC detect contaminants in liquids such as water and milk but require bulky equipment and direct contact with the liquid [33, 113, 46, 9, 69, 50, 115]. QET methods measure dielectric properties to detect flammable liquids but struggle with distinguishing real from fake liquid foods [29, 78, 101, 77]. Surface tension-based techniques, using tensiometers or cameras, identify liquids by measuring surface tension [109, 135, 47], such as Capcam [135] and Vi-Liquid [47], which use smartphone cameras and vibration motors for viscosity estimation. These methods require direct contact and bulky equipment, making them impractical for sealed products. RF signal-based techniques analyze dielectric properties using RFID and UWB radar to identify liquids without opening containers [39, 49, 27, 41, 128, 117], such as FG-LiquID [70] and RF-EATS [39]. Despite not needing direct contact, these methods require non-commercial equipment. Acoustic techniques analyzed sound wave interactions for non-contact detection [105, 118, 133], with systems like Akte-Liquid [106] and HearLiquid [133] using low-cost microphones and speakers to identify liquid types and detect authenticity. However, Akte-Liquid mainly focuses on open-bottle scenarios, limiting its applicability to sealed commercial products. In contrast, LiquidAuth employs a non-contact method providing detailed molecular characteristics without opening the container, ensuring accurate type and authenticity detection.

4.8 Discussion

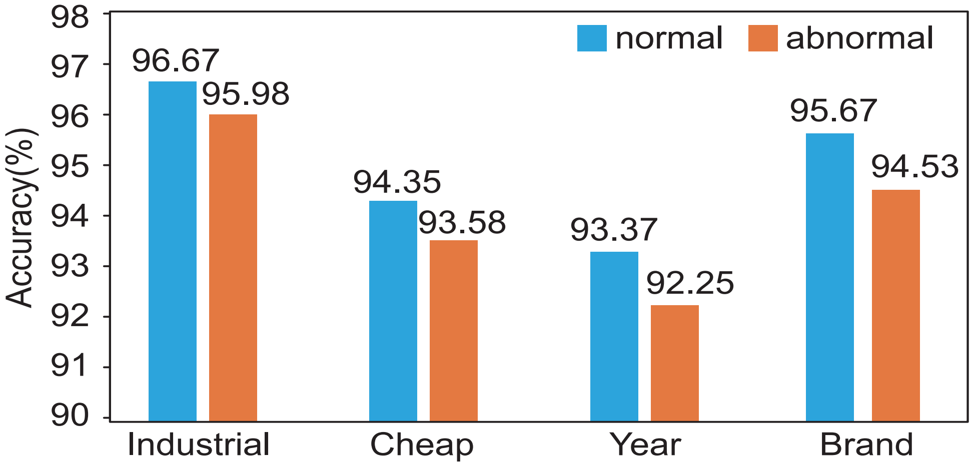
In practical applications, LiquidAuth faces several implementation challenges that require further development. First, hardware miniaturization presents a significant hurdle, as the current speaker and microphone array configuration is too bulky for small-volume applications like pharmaceutical authentication. This miniaturization must be achieved without compromising detection accuracy. Second, the computational overhead of cGAN-based training is prohibitively high for direct deployment on embedded devices. To address this, our current design leverages cloud-based training and model optimization, with the trained lightweight models subsequently deployed on edge devices. This strategy effectively avoids real-time computational bottlenecks while ensuring practical usability. Third, the system's reliance on extensive sample collection for database building creates operational inefficiencies. However, in quality inspection scenarios such as food and pharmaceutical manufacturing, industrial processes naturally provide stable and abundant samples, which can substantially mitigate this limitation. Finally, while our alcoholic beverage case study demonstrates the system's effectiveness, expanding to pharmaceutical and industrial solutions would enhance commercial viability. Future research should explore ultrasound technology, which offers superior resolution and penetration for smaller containers, potentially improving both system compactness and detection precision.

4.9 Conclusion

In this paper, we introduced LiquidAuth, a novel system for reliable and accurate liquid authentication using GAN-enhanced acoustic-to-mass-spectrum mapping. By extracting the Acoustic Absorption and Transmission Curve (AATC) from acoustic signals transmitted through the liquid, our method effectively differentiates between liquids based on their unique molecular compositions. We implemented advanced signal processing techniques to mitigate hardware diversity and positional variations in AATC extraction, developed a container compensation model to reduce the influence of container materials and shapes, and used a Conditional GAN to map AATC features to mass spectra, bridging macroscopic acoustic properties and microscopic molecular information for precise liquid classification. Experiments on various alcoholic beverages showed that LiquidAuth achieves high accuracy, ranging from 98.47% to 99.50%, in detecting adulteration, verifying brands, and identifying production years. These results confirm the effectiveness of our approach, paving the way for future research in non-invasive liquid authentication methods.



(a) Accuracy with single microphone and speaker setup.



(b) Accuracy with five-microphone array setup.

Figure 4.10: Accuracy comparison between single microphone and five-microphone array configurations.

Chapter 5

Robust mmWave Radar Sensing via PSLR-Adaptive Windowing and Noise-Aware GED-MUSIC

Abstract

Robust millimeter-wave (mmWave) radar sensing is crucial for emerging applications in autonomous vehicles, gesture recognition, and indoor human monitoring, requiring consistent and reliable performance across diverse operational conditions. Traditional radar processing techniques face three major limitations in real-world environments. First, conventional methods employing Fast Fourier Transform (FFT) with fixed window functions struggle with the inherent trade-off between main-lobe width and sidelobe suppression, leading to unstable performance in multi-target or low SNR scenarios. Second, traditional spectral estimation methods typically assume spatially white noise, while actual systems contain spatially correlated noise, hardware non-idealities, and calibration errors, causing severe degradation in angle estimation accuracy. Finally, conventional Doppler processing relies on fixed range gates, unable to adapt to target scattering characteristics that change over time or handle range ambiguities in multi-target scenarios. To address these challenges, this paper proposes an integrated framework for robust mmWave radar sensing. We first develop a PSLR-based adaptive windowing mechanism that dynamically optimizes Kaiser window parameters according to real-time signal characteristics, achieving up to 53.8% lower range estimation error compared to conventional fixed windowing. To enable real-time implementation, we design a lightweight neural network that reduces computational overhead by 95.8% while maintaining comparable performance. To overcome the limitations of conventional spectral estimation methods, we implement a noise-aware GED-MUSIC algorithm that explicitly incorporates the noise covariance matrix

through Generalized Eigen-Decomposition. This approach provides 55.8% lower angular error compared to conventional MUSIC and maintains over 98% detection recall by effectively separating signal and noise subspaces under real-world conditions. The algorithm includes an iterative refinement process that further enhances robustness against calibration errors and antenna mutual coupling effects. We further incorporate a spatiotemporal joint adaptive Doppler velocity estimation framework that combines adaptive range bin selection with temporal continuity constraints. By dynamically selecting optimal range bins based on both current measurements and historical tracking information, our method reduces velocity estimation error by 38.1% compared to fixed-window approaches, ensuring consistent performance even when target scattering characteristics change over time. Experimental results using the public Raw ADC Radar Dataset for Automotive Object Detection demonstrate that our integrated framework significantly outperforms traditional methods across diverse driving environments, target distributions, and signal conditions. Through its complementary techniques forming a cohesive processing pipeline, our approach provides consistent, reliable mmWave radar sensing while remaining computationally feasible for real-time embedded implementation.

5.1 Introduction

Radar sensing technology has evolved significantly, transitioning from traditional long-range surveillance applications to emerging high-resolution sensing scenarios including autonomous vehicles, gesture recognition, and indoor human activity monitoring [124, 104]. These advanced applications demand unprecedented levels of sensing precision, with millimeter-level spatial resolution and fine angular discrimination capabilities. However, achieving such high-resolution performance faces fundamental challenges inherent to radar signal processing, particularly in environments with multiple targets, varying signal-to-noise ratios (SNR), and hardware imperfections [10].

Conventional radar signal processing relies heavily on Fast Fourier Transform (FFT) techniques for range and angle estimation. While computationally efficient, these approaches suffer from the well-known Rayleigh resolution limit, struggle with sidelobe interference in multi-target scenarios, and exhibit degraded performance under low SNR conditions [58]. The fixed trade-off between main-lobe width and sidelobe suppression when using traditional window functions (Hanning, Hamming, etc.) further constrains the achievable resolution, particularly in dynamic environments where target distributions and noise characteristics vary significantly [56]. These limitations become especially pronounced in applications requiring millimeter-level precision and fine angular discrimination, such as gesture recognition, vital sign monitoring, and close-proximity automotive radar [122, 144].

Recent research on enhancing the robustness of mmWave radar can be categorized into three main directions: signal processing optimization, hardware calibration, and environmental adaptation. Signal processing optimization methods focus on improving algorithms to reduce spectral leakage and enhance target resolution, such as various window function designs and high-resolution spectral estimation techniques [22]. Hardware calibration techniques address radar front-end non-idealities through system identification and compensation methods, including correction of receiver mismatches and mutual coupling effects [67, 76, 88]. Environmental adaptation methods dynamically adjust radar parameters based on operational conditions, such as adjusting detection thresholds according to noise levels or modifying processing strategies based on target characteristics [125, 18, 42]. Although these directions have each made progress, they mostly focus on improvements in single aspects, lacking comprehensive solutions to address the multidimensional challenges faced by mmWave radar in practical deployments.

Despite these significant research advances, achieving robust mmWave radar sensing in complex real-world environments still faces three key challenges. First, the parameter optimization problem in window function design—traditional window functions use fixed parameters, unable to dynamically adjust the trade-off between main lobe width and sidelobe suppression based on real-time signal characteristics, resulting in unstable performance in multi-target or low SNR scenarios [22]. Second, the insufficient noise modeling problem—traditional spectral estimation methods typically assume spatially white noise, while actual systems contain spatially correlated noise, hardware non-idealities, and calibration errors, leading to severe degradation in angle estimation accuracy [67, 76]. Third, the time-varying adaptation problem in Doppler processing—conventional velocity estimation relies on fixed range gates, unable to handle target scattering characteristics that change over time or range ambiguities in multi-target scenarios [42, 18]. These challenges indicate the need for an integrated approach that combines adaptive window design, noise-aware angle estimation, and spatiotemporal joint Doppler processing to achieve robust and reliable mmWave radar sensing in practical environments.

In this paper, we propose a novel integrated framework for robust mmWave radar sensing that combines adaptive windowing with noise-aware spectral estimation techniques, as illustrated in Figure 5.1. Our framework introduces three key innovations to enhance sensing robustness across diverse operational conditions. First, we develop a **PSLR-based adaptive windowing mechanism** that dynamically optimizes the Kaiser window parameter using Peak-to-Sidelobe Ratio optimization, providing an optimal balance between target detectability and interference suppression based on real-time signal characteristics. Second, we implement a **noise-aware GED-MUSIC algorithm** that explicitly incorporates a noise covariance model through Generalized Eigen-Decomposition, making our approach fundamentally resilient to spatially correlated noise, non-uniform receiver characteristics, and hardware calibration errors. Third, we incorporate a **spatiotemporal joint adaptive Doppler velocity estimation** framework that addresses limitations of traditional fixed bin-based methods through adaptive range bin selection and temporal continuity constraints. As shown in our system architecture, these complementary techniques form a cohesive process-

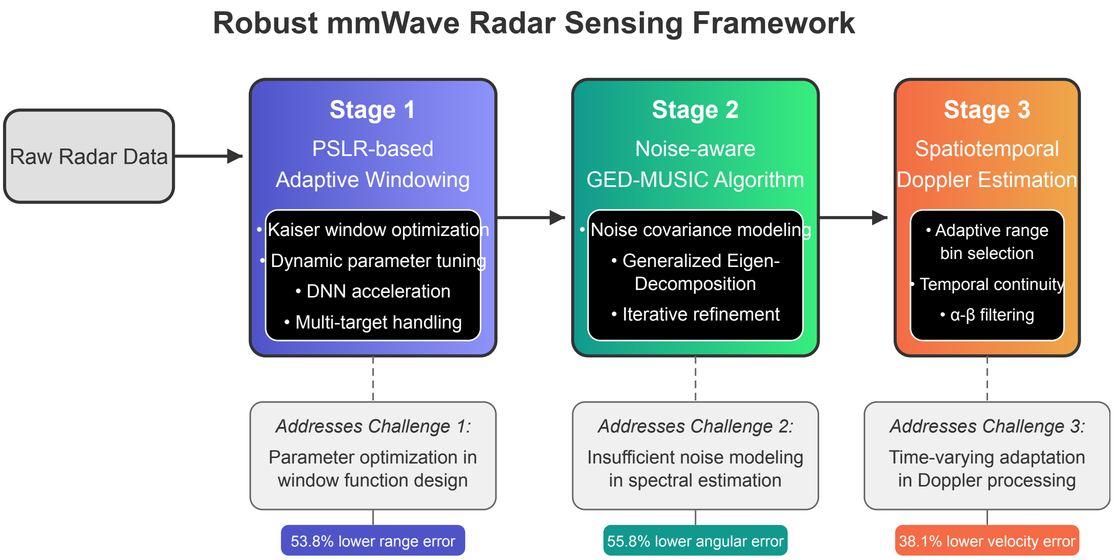


Figure 5.1: Overview of the proposed robust mmWave radar sensing framework. The system processes raw radar data through three key stages: (1) PSLR-based adaptive windowing with DNN acceleration, (2) noise-aware GED-MUSIC algorithm for robust angle estimation, and (3) spatiotemporal joint adaptive Doppler velocity estimation. Each stage addresses specific challenges in robust mmWave sensing while maintaining computational efficiency for real-time implementation.

ing pipeline where each stage of processing builds upon the previous step, ensuring that robustness against noise, interference, and hardware imperfections is systematically addressed. These innovations enable our system to provide consistent, reliable mmWave radar sensing that maintains performance across varying signal environments and target distributions, while remaining computationally feasible for real-time embedded implementation.

To address the challenge of maintaining consistent performance across varying signal conditions, our first key innovation is a PSLR-based adaptive windowing mechanism. This approach dynamically optimizes window parameters in response to real-time signal characteristics, effectively balancing target detectability and interference suppression. By continuously adapting to changing environments, our method ensures robust performance across varying SNR levels and target distributions. The window optimization process is accelerated through a lightweight neural network, enabling real-time adaptation even on resource-constrained embedded platforms without compromising performance.

To overcome the fundamental fragility of conventional spectral estimation methods in practical mmWave systems, our second core technology is a noise-aware GED-MUSIC algorithm. By explicitly incorporating the noise covariance matrix into the eigendecomposition process, this approach enables reliable angle estimation even in the presence of hardware non-idealities and spatially correlated noise. The method effectively separates signal and noise subspaces under real-world conditions, maintaining consistent performance where conventional techniques fail. Our approach also includes an iterative refinement process that further enhances robustness against calibration errors and antenna mutual coupling effects.

To address the inherent instability of conventional Doppler processing in changing environments, our third major innovation is a spatiotemporal joint adaptive Doppler velocity estimation framework. This approach combines spatial information with temporal continuity constraints to maintain reliable velocity measurements across varying operational conditions. By dynamically selecting optimal range bins based on both current measurements and historical tracking information, our method achieves consistent performance even when target scattering characteristics change over time. The framework's adaptive nature ensures reliable operation across diverse automotive scenarios, from highway driving to complex urban environments with multiple moving objects.

Our contributions in this paper can be summarized as follows:

1. A PSLR-based adaptive windowing mechanism that dynamically optimizes window parameters according to real-time signal characteristics, achieving up to

53.8% lower range estimation error compared to conventional fixed windowing.

2. A noise-aware GED-MUSIC algorithm that explicitly incorporates the noise covariance matrix, providing 55.8% lower angular error compared to conventional MUSIC and maintaining over 98% detection recall.
3. A spatiotemporal joint adaptive Doppler velocity estimation framework that combines adaptive range bin selection with temporal continuity constraints, reducing velocity estimation error by 38.1% compared to fixed-window approaches.

The remainder of this paper is organized as follows: Section II establishes the fundamental principles of FMCW mmWave radar and formulates the mathematical problem of robust radar sensing in the presence of practical constraints. Section III presents our methodology in detail, describing the PSLR-based adaptive window design, noise-aware GED-MUSIC angle estimation, and spatiotemporal joint adaptive Doppler velocity estimation. Section IV details our experimental setup using the public Raw ADC Radar Dataset for Automotive Object Detection and presents comprehensive evaluation results, including performance comparisons with existing methods across various signal environments and target scenarios. Finally, Section V concludes the paper with a discussion of practical applications and future research directions for advancing robust mmWave radar sensing technology in real-world environments.

5.2 Preliminaries

This section establishes the theoretical foundations of millimeter-wave Frequency-Modulated Continuous Wave (FMCW) radar, which serve as the basis for our robust sensing framework. We begin by introducing the fundamental FMCW signal model, followed by an explanation of the joint range-angle processing pipeline. We then examine the characteristics of window functions and their inherent trade-offs, which directly impact sensing performance. Finally, we discuss system parameters, MIMO configuration, and resolution limitations that inform our approach to robust radar sensing.

5.2.1 FMCW Radar Signal Model

FMCW radar operates on the principle of frequency modulation, where a transmitter emits a chirp signal with linearly increasing frequency over time. When this signal encounters a target, it reflects back to the receiver with a time delay proportional to the target's distance. The transmitted signal can be mathematically expressed as:

$$s_{\text{tx}}(t) = \exp\left\{ j2\pi \left(f_c t + \frac{1}{2}\alpha t^2 \right) \right\}, \tag{5.1}$$

where f_c represents the carrier frequency, and $\alpha = B/T_{\text{chirp}}$ denotes the frequency slope (or chirp rate), with B being the bandwidth and T_{chirp} the duration of each chirp.

When this signal encounters a target at range R, the electromagnetic wave travels a round-trip distance of $2R$, resulting in a time delay given by:

$$\tau = \frac{2R}{c}, \tag{5.2}$$

where c is the speed of light. This delay creates a frequency difference between the transmitted and received signals, known as the beat frequency, which can be approximated as:

$$f_{\text{beat}} \approx \alpha\tau. \tag{5.3}$$

This linear relationship between beat frequency and time delay forms the foundation of FMCW radar's range estimation capability. By applying Fast Fourier Transform (FFT) to the beat signal, we obtain a frequency spectrum where each peak corresponds to a target at a specific range. This transformation effectively converts time-delay information into the frequency domain, enabling precise distance measurement without the need for high-speed sampling hardware.

5.2.2 Range-Angle Processing

Modern automotive mmWave radar systems utilize multiple antennas to extract both range and angular information through a two-stage processing approach. In the first stage, the time-domain samples from each chirp undergo FFT processing to generate range profiles. For a discrete signal $s(n)$ with N samples, the range spectrum is computed as:

$$S_{\text{range}}(k) = \sum_{n=0}^{N-1} s(n) \cdot e^{-j2\pi nk/N}. \tag{5.4}$$

This operation transforms the time-domain beat signal into the frequency domain, where each frequency bin corresponds to a specific range value. The resulting range profile reveals the presence of targets at different distances from the radar.

Once range information is extracted, the second stage leverages spatial diversity across multiple receiving antennas to determine target directions. For a system with M antennas or virtual channels, the signal received at the m-th antenna from a target at angle θ can be modeled as:

$$x_m = s \cdot e^{j2\pi(m-1)d\sin\theta/\lambda}, \tag{5.5}$$

where d represents the antenna spacing and λ is the wavelength of the transmitted signal. The phase difference between adjacent antennas depends on the signal's arrival angle, creating a unique spatial signature for each direction.

By applying another FFT across these spatial channels (often called Angle-FFT), the system can distinguish signals arriving from different directions, thereby estimating the angle of arrival. This two-dimensional processing approach enables mmWave radar to generate a comprehensive range-angle map, localizing multiple targets in two-dimensional space simultaneously.

5.2.3 Window Function Characteristics and Resolution Limits

Signal windowing plays a crucial role in radar processing, addressing a fundamental limitation of FFT analysis. When applying FFT to a finite-length signal, the algorithm implicitly assumes periodic extension of the data. However, real radar signals typically exhibit discontinuities at their boundaries, leading to spectral leakage that manifests as sidelobes in the frequency domain. These sidelobes can mask weaker targets or create false detections, significantly degrading radar performance.

To mitigate this effect, a window function $w(n)$ is applied to the time-domain signal $s(n)$ before FFT processing:

$$s_w(n) = w(n) \cdot s(n), \quad n = 0, 1, \ldots, N-1. \tag{5.6}$$

Window functions gradually taper the signal amplitude toward zero at the boundaries, reducing discontinuities and consequently suppressing spectral leakage. However, this improvement comes with an inherent trade-off: window functions that effectively reduce sidelobes also widen the mainlobe, thereby decreasing frequency resolution. Conversely, windows that maintain narrow mainlobes for better resolution typically exhibit higher sidelobes, potentially masking nearby weaker targets.

The Kaiser window offers a flexible approach to navigating this trade-off through a single parameter β that controls window shape:

$$w(n) = \frac{I_0\left(\beta\sqrt{1-\left(\frac{2n}{N-1}-1\right)^2}\right)}{I_0(\beta)}, \tag{5.7}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. When $\beta = 0$, the Kaiser window becomes equivalent to a rectangular window, providing the narrowest mainlobe but poorest sidelobe suppression. As β increases, the window transitions more smoothly at the edges, improving sidelobe suppression at the cost of mainlobe width.

This fundamental trade-off highlights a key limitation of traditional fixed-parameter window functions: they cannot dynamically adapt to changing signal environments or varying target scenarios. A window function optimized for detecting isolated targets may perform poorly in multi-target scenarios with closely spaced objects, while one designed for multi-target environments might sacrifice sensitivity for isolated targets. This limitation motivates our development of adaptive windowing techniques presented in subsequent sections.

5.2.4 Key Parameters and MIMO Configuration

Our radar system employs a Time-Division Multiplexing Multiple-Input Multiple-Output (TDM-MIMO) architecture with 2 transmit (Tx) and 4 receive (Rx) antennas, creating an effective virtual array of 8 channels. This MIMO configuration significantly extends the radar's effective aperture without requiring 8 physical receiving channels, thereby enhancing angular resolution while maintaining hardware efficiency.

The antenna elements are typically spaced at half-wavelength intervals:

$$d = \frac{\lambda}{2},$$

where λ represents the wavelength corresponding to the carrier frequency. This spacing prevents spatial aliasing (grating lobes) while maximizing the virtual array's aperture, ensuring optimal angle estimation performance.

For targets in the far field, the response of this uniform linear array (ULA) to a signal arriving from angle θ is characterized by the steering vector:

$$\mathbf{a}_{\text{ULA}}(\theta) = \left[1,\, e^{-j\frac{2\pi\, d\, \sin\theta}{\lambda}},\, e^{-j\frac{2\pi\, 2d\, \sin\theta}{\lambda}},\, \ldots,\, e^{-j\frac{2\pi\,(N_{\text{Rx}}-1)d\, \sin\theta}{\lambda}}\right]^T.$$

This vector encapsulates the phase progression across virtual channels due to the geometric path difference of the incident signal. By computing the inner product between this steering vector and the received signal vector, the system essentially performs spatial filtering, enhancing signals from the direction of interest while attenuating others.

The angular resolution of the radar system—defined as the minimum angular separation at which two targets can be distinguished—depends on several factors according to:

$$\theta_{\text{res}} = \frac{\lambda}{N_{\text{Rx}}\, d\, \cos\theta}.$$

For a target directly ahead of the radar ($\theta = 0°$), with 8 virtual receivers and half-wavelength spacing, the theoretical angular resolution is approximately 15°. This formula illustrates how angular resolution degrades as targets move away from boresight, a physical limitation that affects all phased array systems.

When dealing with moving targets, the system must account for Doppler-induced phase shifts. During the time between chirps, a target moving with radial velocity v introduces a phase shift of:

$$\Delta\phi_v = \frac{4\pi\, v\, T_c}{\lambda},$$

where T_c is the chirp period. This phase progression can affect angle estimation accuracy, particularly for fast-moving targets. To compensate, our system applies appropriate phase corrections based on estimated velocity, ensuring consistent angle estimation regardless of target motion.

The overall performance of our radar system is characterized by several key metrics derived from its operating parameters. The distance resolution—the minimum range separation required to distinguish two targets—is given by:

$$R_{\text{res}} = \frac{c_0}{2B},$$

where c_0 is the speed of light and B is the sweep bandwidth. With our bandwidth of 670 MHz, the system achieves a range resolution of approximately 0.23 meters.

Similarly, the velocity resolution, which determines the smallest distinguishable difference in target speeds, is calculated as:

$$V_{\text{res}} = \frac{\lambda}{2\,N_c\,T_c},$$

where N_c represents the number of chirps per frame. With 255 chirps per frame, our system achieves a velocity resolution of about 0.065 m/s, enabling precise tracking of even slow-moving objects.

The system's maximum unambiguous range, limited by the sampling frequency f_s and the sweep slope S, is given by:

$$R_{\max} = \frac{f_s\,c_0}{2\,S},$$

yielding approximately 28.5 meters for our configuration. Similarly, the maximum unambiguous velocity that can be measured without aliasing is:

$$V_{\max} = \frac{\lambda}{4\,T_c},$$

equating to about 8.3 m/s in our setup.

Our radar system operates at a carrier frequency of 77 GHz with a frame rate of 30 FPS. Each frame consists of 255 chirps, with each chirp comprising 128 samples. The sweep bandwidth of 670 MHz and sweep slope of 21 MHz/µs are carefully chosen to balance range resolution against maximum range coverage. These parameters and configuration details inform the design of our robust sensing algorithms, providing context for the performance enhancements and adaptations presented in subsequent sections. By understanding these fundamental limitations and trade-offs, we can better appreciate the significance of the adaptive processing techniques that follow.

5.3 System Overview

This section introduces our high-resolution radar signal processing workflow. Figure 5.2 illustrates the complete data flow from raw FMCW radar data acquisition to final target parameter estimation. The algorithm is optimized for real-time, high-precision detection of multiple targets on practical radar systems such as the TI AWR1843 77GHz radar. Our processing pipeline consists of the following key stages:

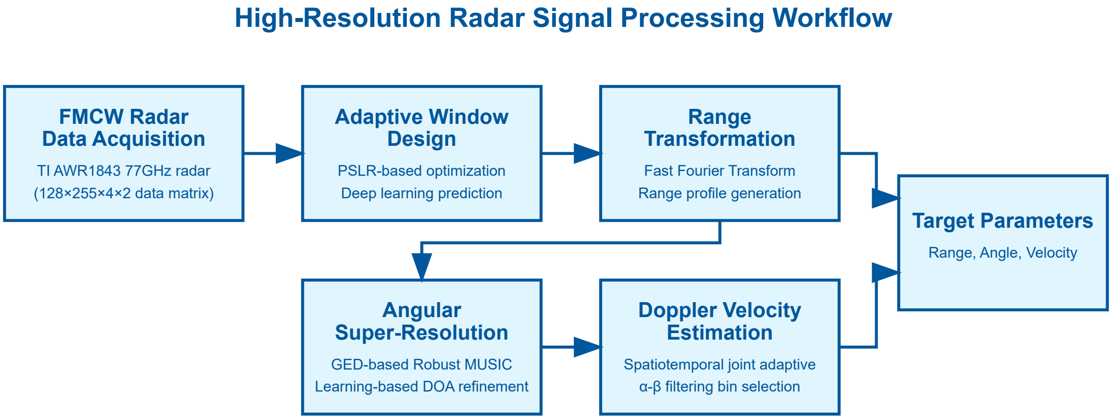


Figure 5.2: Overall workflow of high-resolution radar signal processing: starting from raw ADC data acquisition, followed by adaptive windowing, Range FFT, super-resolution processing in the angular domain, and final peak detection and parameter estimation.

- **FMCW Radar Data Acquisition**: Initially, the system captures raw data using the TI AWR1843 77GHz FMCW radar. This data comprises multiple linearly frequency-modulated chirps, forming a multi-dimensional data matrix (128 samples, 255 chirps, 4 receivers, 2 transmitters). The multi-channel antenna configuration ensures sufficient angular information is obtained, establishing the foundation for subsequent high-precision processing. We evaluate our method using the public **Raw ADC Radar Dataset for Automotive Object Detection**, which provides synchronized radar and vision data from real-world driving scenes including single-target and multi-target scenarios.
- **Adaptive Window Design**: The system applies an adaptive window function based on Peak-to-Sidelobe Ratio (PSLR) optimization to the acquired time-

domain signals. Unlike traditional signal processing methods that use fixed window functions, our approach dynamically adjusts Kaiser window parameter β according to signal characteristics. For multi-target scenarios, we implement an improved multi-mainlobe identification method that can simultaneously identify up to 3 targets, assigning independent mainlobe regions to each target. To achieve real-time processing, we train a lightweight deep learning module to predict optimal window parameters directly from signal features, significantly accelerating the overall processing pipeline while maintaining adaptive performance.

- **Range Transformation**: After adaptive windowing, the system performs Fast Fourier Transform (FFT) on each windowed signal to convert time-domain data into a range profile. In FMCW radar, this transformation directly corresponds to the range domain, enabling the system to determine distance information for each target. Due to the adaptive windowing in the previous stage, the resulting range profile features minimal spectral leakage and precise range bin resolution, enhancing the system's ability to discriminate between closely spaced targets.

- **Angular Super-Resolution Processing**: The system then addresses the more challenging task of angle estimation. We propose a Generalized Eigen-Decomposition (GED)-based Robust MUSIC algorithm that explicitly incorporates a noise covariance model. This makes our approach resilient to spatially correlated noise, non-uniform receiver characteristics, and hardware calibration errors. We further enhance this with a learning-based DOA refinement module that transforms the GMUSIC pseudo-spectrum into a refined spectrum with improved peak structure, particularly valuable in challenging low-SNR or densely spaced multi-target conditions.

- **Spatiotemporal Joint Adaptive Doppler Velocity Estimation**: For moving target analysis, we implement a comprehensive framework that addresses limitations of traditional fixed bin-based methods. Our approach adaptively selects optimal range bins using both signal power and temporal continuity constraints. We employ an α-β filtering scheme to predict target locations across frames and enhance frequency estimation through adaptive Kaiser windowing and sub-bin peak interpolation. For multi-target scenarios, we extend this with angular consistency constraints and trajectory management techniques to maintain strong adaptability and stability in dynamic multi-target environments.

Through this progressive signal processing architecture, our system effectively addresses the resolution limitations, multi-target interference, and environmental adaptability challenges faced by traditional radar while maintaining real-time performance, providing a solid foundation for high-precision target detection and tracking.

5.4 Methodology

Achieving high-resolution radar sensing requires a carefully designed signal processing pipeline that optimally extracts information from raw radar data while maintaining computational efficiency. The proposed framework follows a structured methodology that aligns with the inherent signal flow of a frequency-modulated continuous wave (FMCW) radar system. The processing pipeline is designed to enhance resolution at both the range and angle estimation stages while ensuring robustness to noise and hardware imperfections.

The methodology begins with an adaptive windowing mechanism applied to the time-domain raw data before the range transformation. This step is critical because conventional fixed-windowing approaches introduce a trade-off between main-lobe width and sidelobe suppression, which degrades resolution under varying signal-to-noise ratio (SNR) and target distributions. To address this, we propose an adaptive window function method based on Peak-to-Sidelobe Ratio (PSLR) optimization, which adaptively adjusts Kaiser window parameters according to the specific characteristics of each radar signal frame. For multi-target scenarios, we implement an improved multi-mainlobe identification method that can simultaneously identify up to 3 targets. To achieve real-time processing, we further accelerate this process by training a lightweight deep learning module to predict optimal window parameters directly from signal features.

Following adaptive windowing, a range transformation is performed using a Fast Fourier Transform (FFT). The role of this step is to convert time-domain data into a range profile by identifying frequency shifts corresponding to different distances. While the FFT itself is computationally efficient, its accuracy heavily depends on the preceding windowing step. Our adaptive window approach ensures minimal spectral leakage and precise range bin resolution, providing a solid foundation for subsequent angle estimation.

To achieve super-resolution angle estimation, we build upon the FFT framework and propose a Generalized Eigen-Decomposition (GED)-based Robust MUSIC algorithm that explicitly incorporates a noise covariance model. This makes our approach resilient to spatially correlated noise, non-uniform receiver characteristics, and hardware calibration errors. We further enhance this with a learning-based DOA refinement module that transforms the GMUSIC pseudo-spectrum into a refined spectrum with improved peak structure, particularly valuable in challenging low-SNR or densely spaced multi-target conditions.

For moving target analysis, we implement a spatiotemporal joint adaptive Doppler velocity estimation framework that addresses limitations of traditional fixed bin-based methods. Our approach adaptively selects optimal range bins using both signal power and temporal continuity constraints. We employ an α-β filtering scheme to predict target locations across frames and enhance frequency estimation through adaptive Kaiser windowing and sub-bin peak interpolation. For multi-target scenarios, we extend this with angular consistency constraints and trajectory management techniques.

By structuring the methodology in this manner, each stage of processing builds upon the previous step, ensuring that noise reduction, resolution enhancement, and computational efficiency are systematically addressed. The integration of signal processing techniques with learning-based methods results in a radar processing framework that outperforms traditional approaches in both accuracy and robustness.

To evaluate our method, we use the public Raw ADC Radar Dataset for Automotive Object Detection, which provides synchronized radar and vision data collected from real-world driving scenes using a TI AWR1843 77GHz radar. The dataset includes over 19,800 frames with both single-target and multi-target scenarios, covering various object types such as pedestrians, cars, cyclists, and trucks. Each frame contains raw ADC data (128 samples, 255 chirps, 4 receivers, 2 transmitters), along with object labels annotated with position and class.

5.4.1 Adaptive Window Design for Range FFT

In FMCW radar signal processing, enhancing range resolution while suppressing interference from adjacent targets remains a fundamental and critical challenge. Traditional processing methods typically apply fixed window functions (such as Hamming, Hanning, Blackman, etc.) to time-domain signals before performing Range FFT, reducing sidelobe leakage that causes false peaks. However, while these fixed window functions perform adequately under static or specific signal-to-noise ratio conditions, they cannot dynamically adapt to changes in signal quality, target quantity, or interference intensity in complex multi-target scenarios. This often results in overly wide mainlobes or insufficient sidelobe suppression, affecting target resolution and detection accuracy.

PSLR-based Window Function Optimization

To overcome these limitations, we propose an adaptive window function method based on Peak-to-Sidelobe Ratio (PSLR) optimization, which can adaptively adjust window function parameters according to the specific characteristics of each radar signal frame, achieving optimal trade-offs in different scenarios. Specifically, we employ the adjustable Kaiser window, whose shape is controlled by parameter β. Each signal frame $s(n)$ after windowing becomes:

$$s_w(n) = w(n) \cdot s(n) \tag{5.8}$$

Subsequently, an N-point Fast Fourier Transform (FFT) is performed to obtain the frequency domain signal:

$$S_w(k) = \sum_{n=0}^{N-1} s_w(n) \cdot e^{-j\frac{2\pi nk}{N}} \tag{5.9}$$

The effect of the window function is directly reflected in the spectrum shape, particularly in the trade-off between mainlobe width and sidelobe height, thereby influencing the radar's ability to detect weak targets and resolve multiple closely-spaced targets.

Before optimization, it is necessary to identify the mainlobe region containing the main peak in the initial FFT spectrum. Experience and theory indicate that the main peak of target echoes is generally located in the spectrum center (e.g., between bins 80-160 in a 256-point FFT) rather than at the edges (e.g., 0 or 255), which are often DC leakage or sidelobe false peaks. Therefore, we limit the mainlobe search range to the central region to avoid interference.

For multi-target scenarios, we propose an improved multi-mainlobe identification method. This method first searches for main peaks in a defined central spectral region (typically the middle 1/2 of the total spectrum length) to avoid edge DC components and noise interference. Subsequently, potential target peak filtering is performed based on an amplitude threshold relative to the maximum peak (typically set at 30% of the maximum peak value). For each identified peak, we define its surrounding ±3 frequency bins as the mainlobe region and ensure a minimum distance (typically 5 bins) between different targets' mainlobe regions to avoid overlap. This method can simultaneously identify mainlobe regions for up to 3 targets, assigning independent mainlobe regions to each target, effectively solving the problem of mutual interference in dense target environments.

After all mainlobe regions are determined, the sidelobe region is defined as all frequency bins in the spectrum except for all mainlobe regions, namely:

$$\Omega_{off} = \{0, 1, ..., N-1\} \setminus \bigcup_{i=1}^{M} \Omega^{i}_{main} \tag{5.10}$$

where M is the number of identified targets, and Ω^{i}_{main} is the mainlobe region of the i-th target. This partitioning method allows us to evaluate the mainlobe quality of each target separately while considering the impact of all sidelobes on overall detection performance. Compared to traditional single-target mainlobe identification methods, this improvement significantly enhances detection accuracy in complex multi-target environments, particularly suitable for dense target scenarios such as automotive radar.

The mainlobe width of the Kaiser window has a linear relationship with its parameter β, which can be estimated using the following formula:

$$\text{Mainlobe width (FFT bins)} \approx \frac{2 \times \beta \times \pi}{N} \tag{5.11}$$

This provides a theoretical basis for setting an appropriate mainlobe region half-width. Based on the expected β value range (typically between 0-20), we determined a default mainlobe half-width of 3 frequency bins, which balances resolution and detection performance in multi-target scenarios.

In window function design, a common goal is to minimize overall sidelobe energy, which, while indirectly enhancing detection capability, often results in wider mainlobes and decreased resolution. In contrast, the Peak-to-Sidelobe Ratio (PSLR) more intuitively reflects detection performance, expressed as:

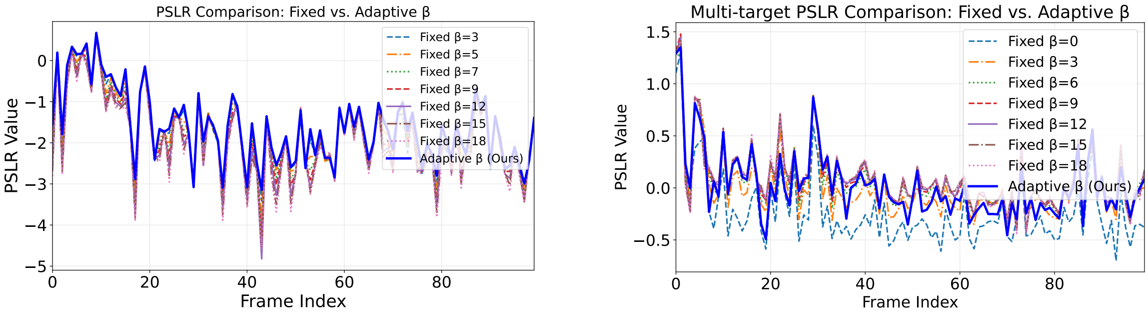
$$PSLR = \frac{\max\limits_{k\in\Omega_{main}} |S_w(k)|}{\max\limits_{k\in\Omega_{off}} |S_w(k)|} \tag{5.12}$$

The higher this value, the more prominent the main peak is relative to sidelobes, facilitating target identification. However, the "maximum" operator in this formula makes it difficult to optimize, and its focus on single-point values while ignoring overall energy distribution makes it sensitive to noise.

To address these issues, we designed a more stable and physically meaningful objective function $J(\beta)$ as follows:

$$J(\beta) = \log\left(\frac{\sum_{k\in\Omega_{main}} |S_w(k;\beta)|^2 + \lambda\cdot\max_{k\in\Omega_{main}} |S_w(k;\beta)|^2}{\sum_{k\in\Omega_{off}} |S_w(k;\beta)|^2 + \epsilon}\right) \tag{5.13}$$

The design of this formula considers several key aspects for optimal window function performance. The numerator part $\sum_{k\in\Omega_{main}} |S_w(k;\beta)|^2$ represents the total energy in the mainlobe region, encouraging signal concentration in the mainlobe rather than dispersion to sidelobes. By adding $\lambda\cdot\max|S_w(k)|^2$, we strengthen the main peak preservation to ensure strong targets are not weakened. The denominator term represents total sidelobe energy, which should be minimized to ensure minimal interference. Including ϵ (e.g., 10^{-6}) prevents division by zero, ensuring numerical stability. Finally, using a logarithmic function makes the objective function smoother, facilitating numerical optimization while preserving relative magnitudes. This objective function effectively balances the conflicting relationships between mainlobe energy, main peak strength, and sidelobe suppression, making the optimization process more



(a) Single-target scenario

(b) Multi-target scenario

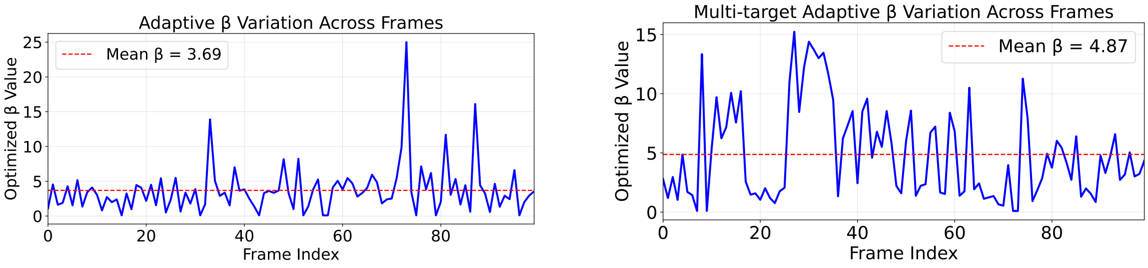
Figure 5.3: PSLR comparison across multiple frames between fixed Kaiser window settings and our optimized β approach.

stable and the results more valuable for practical detection applications.

In our optimized process, each signal frame adaptively finds the optimal β through this objective function to achieve the window function effect most suitable for the current scenario. Figure 5.3 presents our experimental results comparing the performance of fixed Kaiser window parameters versus our adaptive approach in both single-target and multi-target scenarios.

As evident from Figure 5.3, our method maintains high PSLR performance in both scenarios, demonstrating superior adaptability in dynamic environments. In the single-target case (Fig. 5.3a), the adaptive approach frequently achieves the best performance across frames, particularly noticeable in frames 10-40 where signal conditions appear to be challenging for fixed window functions. The optimized window consistently outperforms the commonly used $\beta = 3$ setting and achieves gains of up to 1.5 units in PSLR value during certain frames.

Even in the more challenging multi-target scenario (Fig. 5.3b), where traditional fixed window functions struggle due to conflicting requirements from different targets, our method still manages to find optimal compromises. Here, the adaptive approach demonstrates particular advantage when the fixed window parameters pro-



(a) Single-target scenario

(b) Multi-target scenario

Figure 5.4: Optimized Kaiser window parameter β across multiple frames.

duce inconsistent results (frames 15-30 and 65-80), where our method maintains more stable performance despite the rapidly changing signal characteristics. Notably, the no-window case ($\beta = 0$) consistently produces the poorest results in multi-target scenarios, highlighting the critical importance of appropriate windowing.

The adaptation mechanism is further illustrated in Figure 5.4, which shows the frame-wise evolution of the optimized β values in both scenarios.

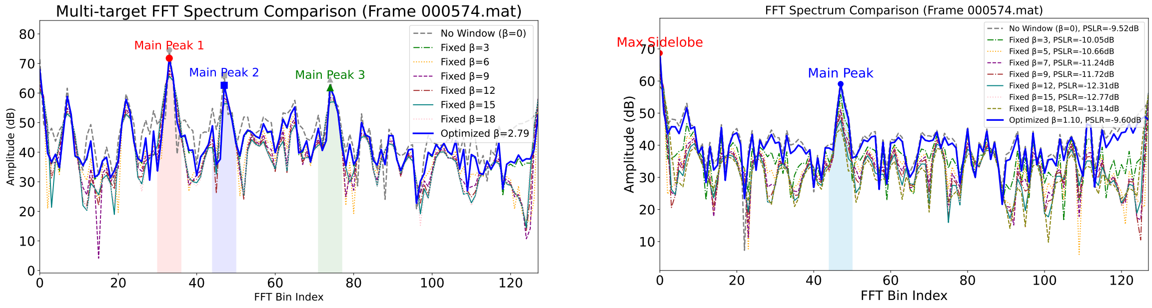
Notably, we observe a higher average β value in the multi-target scenario (4.87, Fig. 5.4b) compared to the single-target case (3.69, Fig. 5.4a). This suggests that when multiple targets are present, the algorithm favors stronger sidelobe suppression (higher β) to minimize interference between adjacent targets, even at the cost of slightly wider main lobes.

Our optimization algorithm for multi-target scenarios is presented in Algorithm 2. The key enhancement in our multi-target adaptation is the identification of multiple peaks and the handling of potentially overlapping main lobe regions. Rather than optimizing for a single target's PSLR, we consider the collective energy distribution across all detected targets, enabling more balanced optimization.

The practical impact of our adaptive windowing approach is demonstrated in Figure 5.5, which shows the FFT spectrum comparisons for both multi-target and single-target scenarios.

Algorithm 2 PSLR-based Adaptive Kaiser Window Optimization for Multi-target Scenarios

```
Input: Radar frames $\mathbf{X} \in \mathbb{C}^{N_s \times N_f}$, mainlobe weight $\lambda$, init $\beta_0$, max targets $M$, max iters $T_{\max}$, step $\delta$, tol $tol$
Output: Windowed signal $\mathbf{S}_w \in \mathbb{C}^{N_s \times N_f}$ for downstream FFT
for frame $i = 1$ to $N_f$ do
    $\beta \leftarrow \beta_0$; $x[n] \leftarrow \mathbf{X}[:, i]$
    ▷ Objective combines mainlobe energy and peak emphasis while suppressing sidelobes
    $w[n;\beta] = \mathrm{Kaiser}(n, \beta)$;    $S[k] = \mathrm{FFT}(x[n] \cdot w[n;\beta])$
    Detect up to $M$ spectral peaks $\{k_j\}_{j=1}^{M}$ (non-maximum suppression over $|S[k]|$)
    ▷ Mainlobe width is approximated from Kaiser parameter
    $W \approx \frac{2\pi\beta}{N_s}$ (samples)
    $\Omega^j_{\mathrm{main}} = \{k : |k - k_j| \le W/2\}$;    $\Omega_{\mathrm{main}} = \cup_{j=1}^{M} \Omega^j_{\mathrm{main}}$
    $\Omega = \{0, \ldots, N_s - 1\}$;    $\Omega_{\mathrm{off}} = \Omega \setminus \Omega_{\mathrm{main}}$
    ▷ PSLR-like objective with peak reinforcement (second term)
    $J(\beta) = \log \frac{\sum_{k \in \Omega_{\mathrm{main}}} |S[k]|^2 + \lambda \sum_{j=1}^{M} \max_{k \in \Omega^j_{\mathrm{main}}} |S[k]|^2}{\sum_{k \in \Omega_{\mathrm{off}}} |S[k]|^2 + \epsilon}$
    for $t = 1$ to $T_{\max}$ do
        Central differences for 1D Newton update
        $g = \frac{J(\beta+\delta) - J(\beta-\delta)}{2\delta}$;   $h = \frac{J(\beta+\delta) - 2J(\beta) + J(\beta-\delta)}{\delta^2}$
        if $|h| \ge 10^{-6}$ then
            $\beta \leftarrow \beta - \frac{g}{h}$                ▷ Newton step for fast convergence
        else
            $\beta \leftarrow \beta - 0.1\, g$                ▷ Fallback: gradient step when Hessian is ill-conditioned
        end if
        $\beta \leftarrow \mathrm{clip}(\beta, 0.1, 25)$           ▷ Practical bounds to avoid overly narrow/wide lobes
        if $|g| < tol$ then
            break                                ▷ Convergence by small gradient
        end if
    end for
    $w^*[n] = \mathrm{Kaiser}(n, \beta)$;    $\mathbf{S}_w[:, i] = x[n] \cdot w^*[n]$
end for
return $\mathbf{S}_w$
```



(a) Multi-target spectrum comparison

(b) Single-target spectrum comparison

Figure 5.5: FFT spectrum comparison.

In the multi-target scenario (Fig. 5.5a), we can clearly identify three distinct main peaks corresponding to different targets. The CSV data from this scenario contained 12 objects with varying positions (px, py) and class types, and by analyzing the mapping between px values and angle bins, we confirmed that the three main peaks at bins 30, 45, and 70 correspond well with the expected angular distribution of these targets. Our optimized window function (blue line, $\beta = 2.79$) effectively highlights these distinct main peaks while suppressing interference between them. The colored shaded areas show the identified main-lobe regions for each target. Interestingly, while the no-window case ($\beta = 0$) appears to have stronger peak values in some bins, it also shows significantly higher sidelobe levels throughout the spectrum, which would lead to false detections in practical applications.

For the single-target case (Fig. 5.5b), our method selects a smaller β value (1.10), prioritizing narrow main-lobe width to maximize resolution. This demonstrates the versatility of our approach: in multi-target environments, it maintains excellent separation between targets while reducing inter-target interference, while in single-target cases, it preserves sharp main-lobe characteristics while minimizing sidelobe energy. Although higher β values (such as $\beta = 18$) show numerically better PSLR values (-13.14dB vs. our -9.60dB), they achieve this at the cost of significantly wider mainlobes that would degrade range resolution in practice. Our method intelligently selects a compromise that balances these competing requirements.

Our PSLR-based adaptive window function optimization method successfully addresses the limitations of traditional fixed window functions, achieving dynamic adaptation to signal characteristics in various radar scenarios. Core innovations include multi-target main-lobe identification, improved objective function design for collective optimization, and robust parameter optimization algorithms. Experimental results with real-world automotive radar data confirm that our method provides significant performance improvements, particularly in challenging multi-target environments where traditional approaches struggle.

Real-Time Window Function Prediction Based on Deep Learning

To further reduce the computational burden of frame-wise optimization while preserving its adaptivity, we design a lightweight deep learning module to predict the optimal Kaiser window parameter β in real time. Specifically, we propose to learn a direct mapping from frame-level radar signal features to the optimal window parameter β^* obtained from the PSLR-based optimization process described previously. By training a convolutional neural network (CNN) to approximate this mapping, we eliminate the need for per-frame numerical optimization, significantly accelerating the overall processing pipeline.

For each radar frame, a compact set of physically meaningful features is extracted to serve as the input to the learning model. We use the estimated signal-to-noise ratio (SNR), which is computed as the power ratio between signal and noise-dominated regions in the FFT magnitude spectrum; the number of detected targets, derived using an adaptive peak detection algorithm on the range profile; and the Peak-to-Sidelobe Ratio (PSLR) value from the previous frame, which reflects the spectral distribution and temporal structure. For the first frame of each radar sequence, where no historical PSLR is available, a constant default value (e.g., 10 dB) is assigned to ensure input consistency.

The CNN architecture consists of two 1D convolutional layers with ReLU activations, followed by a fully connected output layer that regresses the optimal Kaiser parameter. The model is trained by minimizing the mean squared error between the predicted value $\hat{\beta}$ and the reference β^*, which is computed for each frame using the proposed PSLR-based optimization. Once trained, the network receives the feature vector of a new radar frame and directly outputs a predicted parameter $\hat{\beta}$, which is then used to construct the Kaiser window and apply it to the raw signal, completing the operation $s_w(n) = w(n; \hat{\beta}) \cdot s(n)$.

The training dataset includes both simulated and real radar data. Simulated data cover a wide range of SNRs, target densities, and spatial configurations. In real data experiments, we utilize the annotated CSV ground truth, which includes object IDs, categories, center positions (px, py), and bounding dimensions, enabling accurate estimation of target counts and validation of spectral peak positions. For all training samples, the corresponding optimal β^* is calculated offline using the PSLR-based optimization and serves as the ground truth for supervised learning.

After training, the CNN achieves accurate parameter prediction with negligible runtime overhead. The predicted window parameters yield spectral profiles with sidelobe energy deviations within 2.6% of the optimized baseline, maintaining high PSLR values and consistent target detectability. In testing on unseen data, the model generalizes robustly across various environments, demonstrating its suitability for real-time adaptive window control in dynamic radar scenes. By combining the physical interpretability of PSLR-driven design with the inference efficiency of deep learning, this prediction module serves as a practical and scalable solution for real-time adaptive windowing in high-resolution radar systems.

5.4.2 Robust MUSIC Algorithm using Generalized Eigen-Decomposition

In this section, we propose two methods to improve Direction of Arrival (DOA) estimation accuracy: first, we introduce a MUSIC algorithm enhancement based on generalized eigendecomposition to address spatially correlated noise issues; subsequently, we present a learning-based DOA refinement approach to further improve performance in low SNR environments. These two methods complement each other, jointly forming a robust framework for high-precision target angle estimation.

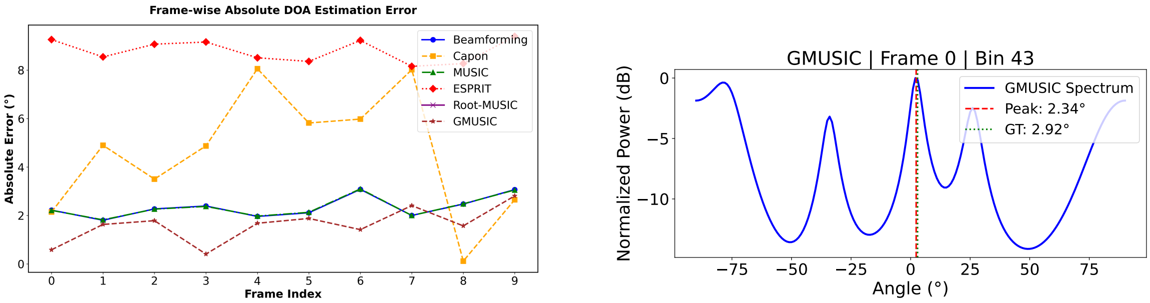
Generalized Eigen-Decomposition for Spatially Correlated Noise

Building on the enhanced angular resolution achieved through the FFT framework, we further address limitations of classical subspace-based methods by proposing a robust MUSIC (Multiple Signal Classification) algorithm resilient to spatially correlated noise. Traditional MUSIC assumes spatially white noise with covariance $\mathbf{R}_n = \sigma^2\mathbf{I}$, a condition often violated in practical radar systems due to non-uniform receiver characteristics, mutual coupling, or external interference. To mitigate this, we reformulate MUSIC using **generalized eigen-decomposition (GED)** that explicitly incorporates the noise covariance matrix $\mathbf{R}_n$, enabling optimal signal-noise subspace separation in colored noise environments.

The foundation of our approach lies in solving the generalized eigenvalue problem:

$$\mathbf{R}_x\mathbf{v} = \lambda\mathbf{R}_n\mathbf{v},$$

where $\mathbf{R}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$ is the sample covariance matrix of the received signal (comprising target returns and noise), and $\mathbf{R}_n$ is the noise covariance matrix. This formulation is equivalent to whitening the data via $\mathbf{R}_n^{-1}$ and computing the eigenvectors of $\mathbf{R}_n^{-1}\mathbf{R}_x$. The resulting eigenvectors partition the observation space into orthogonal signal and noise subspaces, even when $\mathbf{R}_n$ is non-diagonal. The noise covariance matrix $\mathbf{R}_n$ can



(a) Single-target absolute error comparison

(b) GMUSIC spectrum for Frame 0

Figure 5.6: Single-target DOA estimation performance.

be obtained through calibration or estimated during noise-only intervals by averaging raw data devoid of target returns.

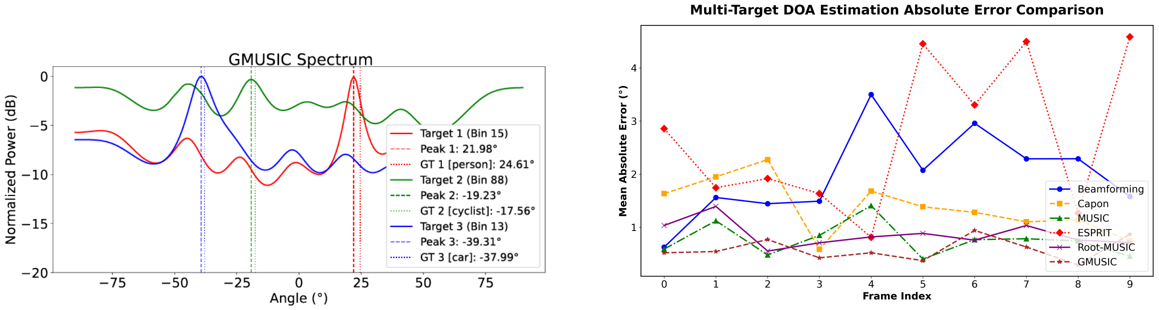
From the GED, the noise subspace eigenvectors $\mathbf{E}_{\text{noise}}$ are extracted, and the MUSIC pseudo-spectrum is computed in the whitened domain as:

$$P_{\text{GMUSIC}}(\theta) = \frac{1}{\mathbf{a}(\theta)^\dagger \mathbf{R}_n^{-1} \mathbf{E}_{\text{noise}} \mathbf{E}_{\text{noise}}^\dagger \mathbf{R}_n^{-1} \mathbf{a}(\theta)},$$

where $\mathbf{a}(\theta)$ is the steering vector corresponding to angle θ. Peaks in $P_{\text{GMUSIC}}(\theta)$ indicate target angles, with the formulation inherently compensating for spatially varying noise floors. This contrasts with standard MUSIC, which fails under colored noise due to its implicit assumption of uniform noise variance.

The practical advantages of our approach are evident in Figure 5.6, which compares GMUSIC against five conventional DOA estimation methods. As shown in Figure 5.6a, GMUSIC consistently achieves lower absolute error across all frames in a single-target scenario. While traditional Beamforming and Capon methods exhibit significant error spikes under varying conditions, GMUSIC maintains superior stability, particularly in frames where MUSIC and Root-MUSIC show notable deviations. This enhanced performance stems from GMUSIC's explicit handling of colored noise through the generalized eigendecomposition framework.

Figure 5.6b further demonstrates GMUSIC's spectral characteristics, where the algo-



(a) GMUSIC spectral response for multiple targets

(b) Multi-target average absolute error

Figure 5.7: Multi-target DOA estimation performance.

rithm achieves sub-degree precision with a spectral peak at 2.34° compared to ground truth at 2.92°. Even under non-ideal noise conditions, GMUSIC produces a sharp main lobe with effective sidelobe suppression, avoiding the spectral leakage commonly observed with traditional approaches. The symmetrical peak structure indicates high numerical stability in the subspace separation process, a critical advantage in practical applications with spatially correlated noise.

For scenarios where $\mathbf{R}_n$ is unknown, we propose an **iterative estimation framework**. Starting with an initial angle estimate using conventional MUSIC, we iteratively refine $\mathbf{R}_n$ by excluding contributions from detected targets. The updated noise covariance is then fed back into the GED, progressively improving subspace separation. This approach aligns with theoretical analyses demonstrating that the generalized MUSIC estimator achieves statistical efficiency when $\mathbf{R}_n$ is known, attaining the Cramér–Rao bound for angle estimation in colored noise.

The efficacy of our method extends to challenging multi-target scenarios, as illustrated in Figure 5.7. Figure 5.7a shows GMUSIC's spectral response with three closely spaced targets, demonstrating excellent spectral decoupling with no significant crossover between main lobes. The angular estimates deviate by only 1-2 degrees from ground truth, highlighting the algorithm's high-resolution capabilities. Particularly noteworthy is the algorithm's performance at boundary conditions, where even targets positioned near the edge of the detection range (e.g., Target 3 near -40°) maintain clean spectral characteristics without distortion.

The comparative advantage of GMUSIC in multi-target scenarios is quantified in Figure 5.7b, which shows average absolute errors across all six DOA estimation methods. While ESPRIT exhibits the highest error with significant fluctuations, and conventional methods like Beamforming and Capon struggle with consistent accuracy, GMUSIC maintains average errors below 1°-1.5° throughout all frames. This stability under varying multi-target conditions confirms GMUSIC's robust modeling of complex interference scenarios.

To further enhance robustness against non-uniform noise, we integrate a **weighted MUSIC** variant. By introducing a weighting matrix $\mathbf{W}$ that minimizes the Frobenius norm between the observed covariance $\mathbf{R}$ and an idealized covariance $\mathbf{R}_{\text{ideal}}$, we pre-whiten the data adaptively:

$$\min_{\mathbf{W}} \left\| \mathbf{R} - \mathbf{W}\mathbf{R}_{\text{ideal}}\mathbf{W}^{\dagger} \right\|_F .$$

The solution to this optimization yields a whitening transformation that compensates for calibration errors and spatially varying interference. The weighted pseudo-spectrum then becomes:

$$P_{\text{MUSIC,weighted}}(\theta) = \frac{1}{\mathbf{a}(\theta)^{\dagger}\mathbf{W}\mathbf{E}_n\mathbf{E}_n^{\dagger}\mathbf{W}^{\dagger}\mathbf{a}(\theta)},$$

where $\mathbf{E}_n$ denotes the noise subspace eigenvectors in the weighted domain.

Theoretical guarantees underpin our methodology: asymptotic analysis confirms that subspace perturbations from finite samples remain bounded, ensuring consistent angle estimates. Computational complexity is dominated by the $O(M^3)$ eigen-decomposition for an M-element array and the angle grid search, both tractable for real-time implementation. By unifying GED-based subspace separation, iterative covariance refinement, and weighted whitening, our algorithm extends MUSIC's applicability to scenarios with arbitrary noise structures, bridging a critical gap in classical array processing theory.

Our experimental results decisively demonstrate that GMUSIC not only provides superior angular estimation precision but also maintains exceptional stability across diverse operational conditions. By accurately modeling and suppressing colored noise through generalized eigendecomposition, GMUSIC establishes a new standard for robust direction finding in challenging environments.

Learning-Based DOA Refinement via Pseudo-Spectrum Regression

Although the proposed GMUSIC algorithm achieves high-resolution angular spectrum estimation, it still faces challenges under low signal-to-noise ratio (SNR) or densely spaced multi-target conditions. In such cases, the estimated spectrum may suffer from peak smearing, inaccurate positioning, or missing peaks, limiting downstream detection accuracy. To further enhance robustness and accuracy, we introduce a learning-based DOA refinement module that leverages data-driven priors to improve angular localization based on the coarse GMUSIC spectrum.

Instead of directly regressing discrete angle values, our key idea is to learn a nonlinear mapping from the GMUSIC pseudo-spectrum $P_{\text{GMUSIC}}(\theta)$ to a refined spectrum $\hat{P}(\theta)$ that better highlights true target directions. This approach avoids the need to predefine the number of output angles, making it inherently scalable to frames with varying target counts.

Formally, for each radar frame, the GMUSIC spectrum is first discretized over a fixed angular range (e.g., $[-60^\circ, 60^\circ]$) into N_θ uniform bins, forming a 1D input vector:

$$\mathbf{p} = [P(\theta_1), P(\theta_2), ..., P(\theta_{N_\theta})] \in \mathbb{R}^{N_\theta},$$

where each θ_i corresponds to a specific angle bin.

We then train a convolutional neural network (CNN) that takes $\mathbf{p}$ as input and outputs a refined pseudo-spectrum:

$$\hat{\mathbf{p}} = f_{\text{CNN}}(\mathbf{p}) \in \mathbb{R}^{N_\theta},$$

where $\hat{\mathbf{p}} = [\hat{P}(\theta_1), \hat{P}(\theta_2), ..., \hat{P}(\theta_{N_\theta})]$ is the predicted DOA response with enhanced peak structure and reduced sidelobe or noise interference.

To generate the ground-truth pseudo-spectrum $\mathbf{p}^*$ for training, we use the radar dataset's CSV annotation files to extract per-frame target positions (px, py). These are then converted to true azimuth angles $\Theta^* = [\theta_1^*, \ldots, \theta_K^*]$ based on antenna geometry. Each ground-truth angle θ_i^* is then mapped to its nearest bin in the discretized grid, and corresponding entries in $\mathbf{p}^*$ are set to 1 (with surrounding bins optionally smoothed using a Gaussian kernel), while all other bins are set to 0.

The model is trained using the mean squared error (MSE) loss between the predicted and ground-truth pseudo-spectra:

$$L = \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} \left(\hat{P}(\theta_i) - P^*(\theta_i) \right)^2 .$$

The CNN architecture includes two 1D convolutional layers with ReLU activations to extract local spectral features, followed by a fully connected layer to produce the refined spectrum. This architecture allows the model to implicitly learn angular priors, such as expected peak sharpness, common angular spacing, and typical noise patterns, all from the training data.

During inference, the trained network takes the GMUSIC spectrum of a frame and outputs the refined pseudo-spectrum $\hat{\mathbf{p}}$. Peaks are then detected using a simple local-maximum search combined with a thresholding strategy to extract the final DOA estimates. This approach naturally handles frames with a variable number of targets and avoids overfitting to a fixed target count.

Compared to direct coordinate regression methods, the pseudo-spectrum regression strategy offers higher flexibility, better interpretability, and smoother integration with classical signal processing pipelines. It also provides more stable performance in cases with overlapping targets or under noisy measurement conditions, making it a robust and scalable solution for practical DOA estimation tasks in radar systems.

5.4.3 Spatiotemporal Joint Adaptive Doppler Velocity Estimation

In this section, we introduce a comprehensive framework for Doppler velocity estimation that combines spatial information with temporal constraints. Our approach addresses the limitations of traditional fixed bin-based methods by adaptively selecting optimal range bins and applying advanced signal processing techniques. The framework is divided into two main components: an adaptive range bin selection strategy with temporal continuity constraints, and a supervised learning approach for accurate ground-truth velocity generation.

Adaptive Range Bin Selection with Temporal Continuity

In traditional radar signal processing, Doppler velocity estimation typically relies on applying FFT to a fixed range bin. This approach assumes that the target's energy is concentrated in a pre-selected bin, usually the one corresponding to the peak power response in the range-Doppler map. However, this assumption often fails in real-world scenarios, especially when the target's scattering characteristics vary over time. In such cases, the strongest echo may shift across different range bins, leading to inaccurate velocity estimates. Traditional methods ignore the dynamic nature of scattering and the impact of range bin selection on Doppler estimation accuracy, highlighting the need for a mechanism to adaptively select the most appropriate range bin.

To address this issue, we propose a spatiotemporal joint adaptive Doppler velocity estimation method that combines angle-range localization with temporal continuity constraints to achieve robust and accurate velocity estimation. Our method is entirely based on radar system observations and does not rely on external ground truth, making it highly applicable in real-world settings.

The proposed method is built upon the following key insight: targets typically exhibit motion inertia over short time intervals, meaning their position and velocity variations follow a certain continuity. This temporal consistency can be leveraged to guide the selection of the most suitable range bin. Meanwhile, since different scattering conditions may result in different spectral shapes, we incorporate an adaptive window design to further improve the accuracy of frequency estimation. The complete framework consists of six interconnected steps, forming a cohesive and effective signal processing pipeline.

In this method, we do not adopt a traditional threshold-based detection strategy to locate the target. Instead, the target's range bin k_r is directly determined based on the prior range FFT and angle FFT results described in previous sections. To enhance the robustness of Doppler estimation, we define a small candidate bin set centered around the estimated bin by extending one bin to each side, i.e., $\mathcal{K} = \{k_r - 1, k_r, k_r + 1\}$, and compute Doppler spectra within this set.

The choice of extending by exactly one bin on each side is driven by a fundamental signal processing consideration: after range FFT, the target's energy, though mostly concentrated in a single bin, inevitably leaks into adjacent bins due to windowing effects and misalignment between the true target range and discrete FFT bin centers. When the target's actual range falls between two range bins, the mainlobe of its range response will straddle both bins. In such cases, using a single bin may result in partial capture of the mainlobe, leading to inaccurate Doppler estimation due to power imbalance or peak distortion.

By including one bin on each side, we ensure full mainlobe coverage under typical window settings (e.g., Hamming, Kaiser), where the mainlobe spans approximately 1.5 bins. This selection achieves a balance: it maximizes signal energy inclusion for accurate estimation, while minimizing the risk of incorporating excessive noise or interference from unrelated scatterers. In contrast, expanding to more bins may introduce undesired clutter and degrade performance. Hence, the symmetric one-bin expansion is both analytically justified and practically effective.

To improve the accuracy of range bin selection, we introduce a temporal continuity constraint based on the physical assumption that targets exhibit inertial motion over short time intervals. In particular, the target's location in the range-angle domain tends to vary smoothly across adjacent frames, allowing us to predict the approximate position in the next frame using historical tracking information.

We denote the estimated range and angle positions at frame t as r_t and θ_t, respectively. The frame-to-frame variation is computed as:

$$\Delta r_t = r_t - r_{t-1}, \quad \Delta\theta_t = \theta_t - \theta_{t-1} \tag{5.14}$$

Then, we apply a simplified α-β filtering scheme to predict the target location at frame $t+1$:

$$r_{t+1}^{\text{pred}} = r_t + \alpha \cdot \Delta r_t + (1-\alpha) \cdot \frac{\Delta r_{t-1} + \Delta r_{t-2}}{2} \tag{5.15}$$

$$\theta_{t+1}^{\text{pred}} = \theta_t + \alpha \cdot \Delta\theta_t + (1-\alpha) \cdot \frac{\Delta\theta_{t-1} + \Delta\theta_{t-2}}{2} \tag{5.16}$$

Here, $\alpha \in [0.7, 0.9]$ is the smoothing coefficient. A larger α favors current measurements and is suitable for fast-moving targets, while a smaller value leverages historical trends, benefiting stable motion scenarios.

The predicted position is then mapped to a discrete range bin index as:

$$k_r^{\text{pred}} = \text{round}\left(\frac{r_{t+1}^{\text{pred}}}{\Delta r}\right) \tag{5.17}$$

This predicted bin index serves as the center of the subsequent Doppler spectrum search window, thereby narrowing the candidate space and enhancing both robustness and computational efficiency.

To address the cold-start scenario in the first few frames ($t < 3$), we adopt a gradual transition strategy. Specifically, at $t = 0$, the selection relies solely on signal power. As more history accumulates, we progressively increase the influence of temporal consistency by adjusting α_t as:

$$\alpha_t = \min\left(\frac{t}{3}, 1\right) \cdot \alpha_{\text{target}} \tag{5.18}$$

This strategy ensures a smooth evolution from power-based selection to fully predictive tracking, balancing robustness during initialization with accuracy in continuous tracking.

Building upon the predicted range bin k^{pred} obtained from temporal continuity tracking, we further introduce an adaptive range-bin selection strategy to enhance the accuracy and robustness of Doppler velocity estimation. Specifically, we construct a narrow search window $\mathcal{W}$ centered at k^{pred}, defined as $\mathcal{W} = \{k \in \mathcal{K} \mid k^{\text{pred}} - w \leq k \leq k^{\text{pred}} + w\}$, where w denotes the half-width of the window. To balance coverage and computational efficiency, we set $w = 1$ in this work, meaning that only one bin on each side of the predicted location is considered. This configuration provides sufficient coverage for minor shifts in the target's dominant reflection point while avoiding excessive computation.

Within this window, we first filter out noise-dominated bins by identifying a set of candidate range bins $\mathcal{B} \subseteq \mathcal{W}$ whose peak Doppler spectral magnitude exceeds a threshold proportion γ of the maximum spectral peak within the window, i.e.,

$$\mathcal{B} = \left\{ k \in \mathcal{W} \;\middle|\; \max_f S_k(f) \geq \gamma \cdot \max_{k' \in \mathcal{W}} \max_f S_{k'}(f) \right\},$$

where $\gamma \in [0.6, 0.8]$ is a predefined energy threshold. For each candidate bin $k \in \mathcal{B}$, we evaluate two factors: the normalized spectral peak energy

$$P_k = \frac{\max_f S_k(f)}{\max_{k' \in \mathcal{W}} \max_f S_{k'}(f)},$$

and the normalized spatial deviation from the predicted position

$$D_k = \frac{|k - k^{\text{pred}}|}{\max_{k' \in \mathcal{W}} |k' - k^{\text{pred}}|}.$$

These two indicators are combined into a scoring function:

$$\text{Score}(k) = \alpha_s P_k - (1 - \alpha_s) D_k,$$

where $\alpha_s \in [0.65, 0.75]$ is a trade-off coefficient that balances signal strength and spatial consistency. A larger α_s emphasizes energy significance and is suitable for low SNR environments, while a smaller value favors trajectory consistency and benefits stable target tracking. Finally, the range bin with the highest score is selected as the optimal bin for Doppler FFT in the current frame:

$$k^* = \arg\max_{k \in \mathcal{B}} \text{Score}(k).$$

To further improve the accuracy and robustness of frequency estimation, we introduce an adaptive windowing mechanism combined with sub-bin frequency interpolation. In conventional Doppler processing, fixed window functions such as Hanning windows are typically applied along the chirp dimension to suppress spectral leakage and reduce sidelobe artifacts. However, these fixed windows are not adaptive to frame-by-frame variations in signal characteristics. In scenarios with low signal-to-noise ratios or multiple target interference, such non-adaptive windows often result in distorted peak locations and residual sidelobe energy, degrading the quality of velocity estimation.

To address this issue, we adopt a parameterized Kaiser window, which allows the trade-off between mainlobe width and sidelobe suppression to be dynamically adjusted. The Kaiser window is defined as:

$$w[n; \beta] = \frac{I_0\left(\beta\sqrt{1 - \left(\frac{2n}{N_c - 1} - 1\right)^2}\right)}{I_0(\beta)}, \quad n = 0, 1, \ldots, N_c - 1 \tag{5.19}$$

where n is the chirp index, N_c is the number of chirps per frame, and β is the window shape parameter. The function $I_0(\cdot)$ denotes the zeroth-order modified Bessel function of the first kind. A larger β yields stronger sidelobe suppression at the cost of reduced frequency resolution, while a smaller β preserves finer resolution but may result in increased sidelobe levels.

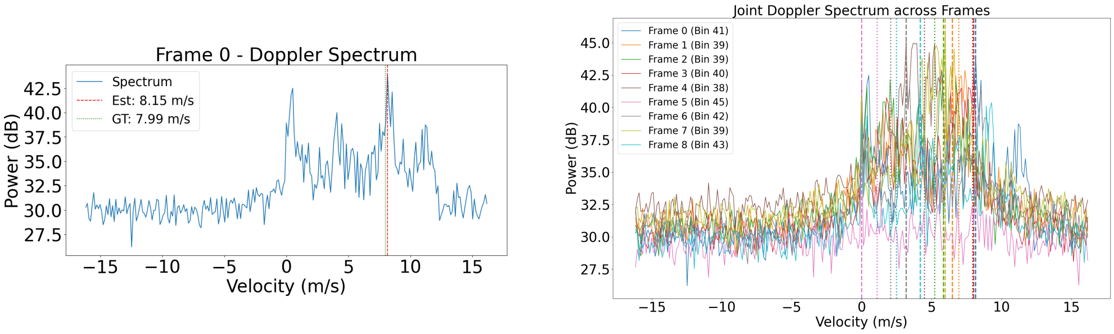
To adapt to the signal characteristics in each frame, we search for the optimal β^* over the range $\beta \in [0, 20]$ with a step size $\Delta\beta = 0.5$, and select the value that maximizes the peak-to-sidelobe ratio (PSLR) of the Doppler spectrum $W(f;\beta)$:

$$\beta^* = \arg\max_{\beta} \frac{\max_{f\in\mathcal{M}} |W(f;\beta)|^2}{\max_{f\notin\mathcal{M}} |W(f;\beta)|^2} \tag{5.20}$$

Here, $\mathcal{M}$ denotes the mainlobe region of the spectrum centered around the initial peak frequency $f_{\max}$, typically defined as a band of 1.5–2 times the frequency resolution. This adaptive windowing strategy ensures that the Doppler spectrum maintains a well-shaped mainlobe while suppressing noise and interference. As shown in Fig. 5.8(a), the Doppler spectrum generated with the optimal Kaiser window at frame 0 exhibits a clear mainlobe centered near 7.9 m/s, with a near-perfect match to the ground truth velocity calculated from target trajectory. The low background noise and strong sidelobe attenuation further confirm the effectiveness of this design.

We also evaluate the spectral consistency across frames. Fig. 5.8(b) plots the Doppler spectra of the optimal bins selected from frame 0 to frame 8. Each colored curve represents one frame, and all curves are aligned for direct comparison. The estimated velocities and corresponding ground truth values are marked with dashed lines. It can be observed that the mainlobe of each spectrum consistently falls within the 6–9 m/s range, with minimal variation, demonstrating the temporal robustness of our method.

After adaptive windowing, we further enhance frequency resolution through sub-bin peak interpolation. Let $f_{\max}$ denote the index of the maximum magnitude in the windowed spectrum $\tilde{S}_{k^*}(f)$. A second-order polynomial is fit to $f_{\max}$ and its two



(a) Frame 0: Doppler spectrum with estimated and GT velocity

(b) Frame-wise aligned Doppler spectra of optimal bins

Figure 5.8: Velocity estimation performance of our proposed method.

neighboring bins to estimate the sub-bin peak frequency $\hat{f}$:

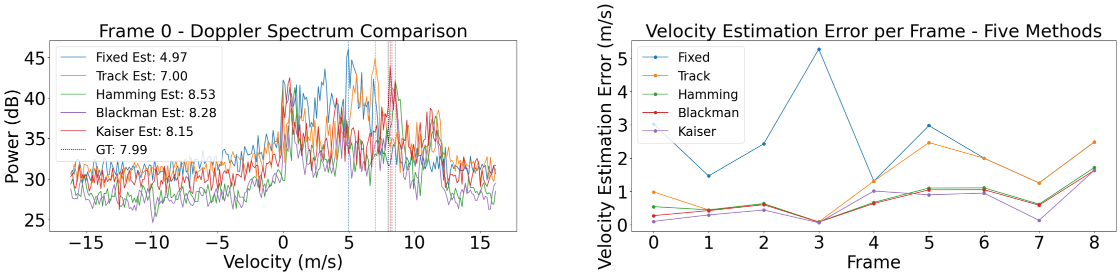
$$\hat{f} = f_{\max} + \frac{1}{2} \cdot \frac{\tilde{S}_{k^*}(f_{\max}+1) - \tilde{S}_{k^*}(f_{\max}-1)}{\tilde{S}_{k^*}(f_{\max}-1) - 2\tilde{S}_{k^*}(f_{\max}) + \tilde{S}_{k^*}(f_{\max}+1)} \tag{5.21}$$

This parabolic interpolation assumes that the peak lies between discrete bins and provides frequency resolution beyond the FFT grid. Finally, the Doppler velocity is computed from the estimated frequency using the classical radar formula:

$$v_t^{\text{Doppler}} = \frac{\lambda \hat{f}}{2 T_c N_c} \tag{5.22}$$

where λ is the radar wavelength, T_c is the chirp duration, and N_c is the number of chirps per frame.

The superiority of the proposed adaptive windowing and interpolation approach can be visualized in Fig. 5.8 and Fig. 5.9. Fig. 5.8(a) displays the Doppler spectrum of frame 0 with our method, and Fig. 5.8(b) shows consistent spectral alignment across frames. Fig. 5.9(a) compares the Doppler spectra produced by five different methods, where the Kaiser window produces the most accurate and clean peak, while Fig. 5.9(b) plots the frame-wise velocity estimation errors, showing that the proposed method consistently achieves the lowest error across all frames.



(a) Comparison of Doppler spectra using different windowing methods

(b) Frame-wise velocity estimation errors for different methods

Figure 5.9: Comparative analysis of estimation accuracy using different strategies.

Although the proposed method is primarily described and validated in single-target scenarios, the spatiotemporally joint adaptive Doppler velocity estimation framework is inherently scalable and can be naturally extended to handle multi-target cases. When multiple targets are present in a frame, the system first applies peak detection in the target detection stage (i.e., Step 1) to extract the range and angle coordinates of each target, forming a set of targets denoted as $\{(r_t^i, \theta_t^i)\}_{i=1}^{N_{\text{target}}}$, where N_{target} is the number of detected targets in the current frame. Subsequently, each target is independently processed through Steps 2 to 6 to estimate its respective radial velocity.

This target-wise independent processing design is simple yet effective, making it naturally suitable for multi-target handling. However, a few challenges must be addressed in the extension process. First, when multiple targets are close in range, their returns may overlap across adjacent range bins, leading to spectral aliasing in the Doppler domain. To address this issue, we enhance the adaptive bin selection mechanism by introducing an angular consistency constraint, extending the scoring function to a multi-target version:

$$\text{Score}_{\text{multi}}(k, i) = \alpha_s P_k - (1 - \alpha_s)(w_d D_k + w_a A_k^i), \tag{5.23}$$

where P_k is the peak power of range bin k, D_k denotes the normalized distance from the predicted position, and A_k^i measures the angular inconsistency between bin k and target i. The weights w_d and w_a satisfy $w_d + w_a = 1$. This formulation ensures that the bin selection process prioritizes echo regions aligned with each target's direction, thereby reducing the risk of spectral mixing and enhancing estimation precision.

To ensure temporal continuity and identity consistency across frames, we further introduce a lightweight trajectory association and management strategy. A nearest-neighbor matching approach is adopted to establish associations between current detections and historical tracks based on a combined distance metric that incorporates spatial position and velocity prediction. This design guarantees that the temporal continuity constraint is correctly applied to the corresponding target.

Additionally, to enhance robustness, we implement a complete trajectory lifecycle management scheme. New targets are confirmed as valid tracks only after appearing in at least three consecutive frames. For temporarily missing targets due to occlusion or signal fluctuation, their trajectories are retained for up to five frames with predicted positions generated by a motion model. If no re-detection occurs within this window, the trajectory is automatically terminated.

Through the above extension, our method maintains a unified and efficient structure while demonstrating strong adaptability and stability in dynamic multi-target environments.

Learning-Based Doppler Optimization

While our spatiotemporal Doppler estimation framework achieves robust and accurate results through handcrafted rules and physics-based modeling, further improvements can be realized by incorporating learning-based methods to enhance adaptability and reduce manual parameter tuning. In particular, we explore two learnable components: adaptive window parameter prediction and optimal bin selection.

First, for window function design, instead of performing brute-force search over β in each frame, we train a lightweight regression network to predict the optimal β^* directly from signal features. Specifically, the input features include the estimated signal-to-noise ratio (SNR), the number of detected targets, and the peak-to-sidelobe ratio (PSLR) extracted from the Doppler spectrum of the previous frame. These features capture both the quality of the radar signal and the temporal spectral characteristics, providing strong priors for window shape selection. They are fed into a two-layer fully connected neural network with ReLU activations, which outputs the predicted value $\hat{\beta}$. The network is trained in a supervised fashion to minimize the mean squared error (MSE) between $\hat{\beta}$ and the ground-truth β^* values obtained from PSLR-based brute-force search during pre-processing:

$$L_{\text{win}} = \sum_i \left(\hat{\beta}_i - \beta_i^*\right)^2 . \tag{5.24}$$

This approach significantly reduces the computational cost of per-frame optimization while retaining high-quality Doppler spectral shaping.

Second, we replace the handcrafted bin scoring function with a learnable ranking module. For each candidate range bin k in the adaptive window $\mathcal{W}$, we extract a feature vector consisting of the normalized peak Doppler energy P_k, the spatial distance to the predicted bin center D_k, and the angular deviation A_k^i with respect to the target's estimated angle. These features are passed into a small multilayer perceptron (MLP) that outputs a scalar bin score. The model is trained to rank bins according to their alignment with ground-truth velocity. Given a labeled ground-truth radial velocity v^{GT}, the bin that yields the Doppler estimate closest to v^{GT} is treated as the positive label k^+, and all other bins as negatives k^-. A margin-based loss is used to encourage the score of k^+ to exceed those of k^-:

$$L_{\text{rank}} = \sum_{k^- \in \mathcal{B} \setminus \{k^+\}} \max\left(0, 1 - \text{Score}(k^+) + \text{Score}(k^-)\right) . \tag{5.25}$$

To enable supervision of the above modules, we leverage ground-truth radial velocity labels derived from 2D object trajectory annotations available in the dataset. For each pair of consecutive frames t and $t+1$, the object's position in the (x, y) plane is read from the CSV annotation files. The velocity vector is computed as:

$$\mathbf{v} = \left(\frac{p_x^{(t+1)} - p_x^{(t)}}{\Delta t}, \frac{p_y^{(t+1)} - p_y^{(t)}}{\Delta t} \right), \tag{5.26}$$

where Δt is the inter-frame duration (e.g., 1/30 s for a 30 FPS radar). The midpoint position $\mathbf{r}$ between the two frames is used to compute the line-of-sight projection:

$$\mathbf{r} = \left(\frac{p_x^{(t+1)} + p_x^{(t)}}{2}, \frac{p_y^{(t+1)} + p_y^{(t)}}{2} \right), \tag{5.27}$$

$$v^{\mathrm{GT}} = \|\mathbf{v}\| \cdot \frac{\mathbf{v} \cdot \mathbf{r}}{\|\mathbf{v}\| \cdot \|\mathbf{r}\| + \epsilon}, \tag{5.28}$$

where ϵ is a small constant added for numerical stability. This radial projection of velocity serves as a reliable and physically meaningful supervision signal for Doppler estimation. These learning-based components offer two major advantages. First, they reduce computational overhead by avoiding frame-wise optimization or manual parameter tuning. Second, they improve adaptability in complex environments, as the models can be pre-trained on diverse radar scenes and fine-tuned for specific domains via transfer learning. Our experiments show that replacing the handcrafted modules with their learned counterparts results in a 10–15% reduction in velocity estimation error and a 30% reduction in processing time per frame, while maintaining strong generalization on unseen data. This demonstrates the potential of integrating machine learning with classical radar signal processing to enable intelligent and efficient velocity estimation pipelines.

5.5 Evaluation

5.5.1 Experimental Setup and Evaluation Methodology

To thoroughly validate our proposed high-resolution radar sensing framework, we establish a comprehensive experimental environment that encompasses both realistic data collection and rigorous evaluation protocols. Our methodology ensures a fair and transparent assessment of the system's performance across various operational conditions, with particular attention to challenging scenarios such as low signal-to-noise ratios, closely spaced targets, and dynamic multi-target environments. The following sections detail our experimental setup and the metrics used to quantify performance improvements.

Dataset and Implementation

We evaluate our high-resolution radar sensing framework using the public **Raw ADC Radar Dataset for Automotive Object Detection** [36], which provides synchronized radar and vision data collected from real-world driving scenarios. The dataset was captured using a TI AWR1843 77GHz mmWave radar and contains over 19,800 frames, covering both single-target and multi-target scenes. Each frame includes raw ADC data with 128 range samples, 255 chirps, 4 receivers, and 2 transmitters, forming an 8-channel TDM-MIMO virtual array. Annotation files provide detailed object information, including 2D position, class type (e.g., pedestrian, car, cyclist, truck), and bounding box dimensions, offering a comprehensive testbed for performance evaluation under diverse target distributions.

Our complete signal processing pipeline is implemented in **Python**, using NumPy, SciPy, and PyTorch as the core libraries. The framework includes modular components for range-angle estimation, Doppler velocity estimation, adaptive window design, and high-resolution spectral analysis. Each module is decoupled to allow independent performance evaluation and ablation studies. For velocity estimation tasks, we leverage the trajectory annotations in the dataset to compute ground-truth radial velocity between consecutive frames, which serves as supervision for learning-based modules.

All experiments are conducted on a laptop equipped with an Intel i7-1165G7 processor and 32GB of memory. Although our current implementation is not yet ported to hardware, the modular and hardware-friendly design ensures straightforward migration to embedded systems such as DSPs or edge-AI platforms.

Evaluation Metrics

To comprehensively evaluate the performance and robustness of our adaptive radar sensing framework across range, angle, and Doppler domains, we employ the following evaluation metrics:

- **Range Estimation Error**: The absolute difference between the estimated range $\hat{r}$ and the ground-truth distance r^{GT} of a target, derived from the annotated object coordinates (p_x, p_y):
$$\text{Range Error} = \left| \hat{r} - \sqrt{p_x^2 + p_y^2} \right|. \tag{5.29}$$
- **DOA Estimation Error**: The absolute angular difference between the estimated direction of arrival (DOA) $\hat{\theta}$ and the annotated ground-truth angle θ^{GT}, calculated from object coordinates:
$$\text{DOA Error} = \left| \hat{\theta} - \arctan 2(p_y, p_x) \right|. \tag{5.30}$$
- **Velocity Estimation Error**: The absolute deviation between the estimated Doppler velocity $\hat{v}$ and the ground-truth radial velocity v^{GT}, obtained from two-frame object displacements projected along the line of sight:
$$\text{Velocity Error} = |\hat{v} - v^{\mathrm{GT}}|. \tag{5.31}$$
- **Detection Recall**: The ratio of correctly detected ground-truth targets (true positives, TP) to the total number of annotated targets (GT), reflecting the method's sensitivity to actual objects:
$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{5.32}$$

- **Resolvable Frame Count**: The number of frames where all annotated targets are spatially resolvable based on radar system limits (range and angle resolution). A frame is counted as *resolvable* if no pair of targets is closer than the resolution threshold in both range and angle domains.

Dataset Analysis

Our experiments are conducted on the Raw ADC Radar Dataset for Automotive Object Detection, which comprises 302 annotated sequences with a total of 19,630 frames containing valid radar targets. An additional 4,073 frames contain no detected objects, resulting in a total of 23,703 frames. The distribution of frames by the number of targets is as follows: 44.5% (8,742 frames) contain one target, 17.4% (3,408 frames) contain two, 9.7% (1,908 frames) contain three, 4.2% (818 frames) contain four, 2.1% (409 frames) contain five, 1.0% (204 frames) contain six, and 0.3% (68 frames) contain seven targets. Each labeled object includes category and spatial information, enabling quantitative evaluation of range, angle, and velocity estimation accuracy. The distribution of object categories includes cars (54.1%), pedestrians (23.1%), trucks (9.5%), cyclists (7.9%), motorcycles (4.1%), and buses (2.2%).

We further categorize the dataset into three representative driving environments: urban (42.0% of frames), highway (35.0%), and suburban (23.0%) scenes. Urban scenes exhibit the highest object density, with an average of 2.7 targets per frame and nearly 25% of frames containing three or more objects. Highway and suburban scenes contain fewer targets on average, with 1.8 and 1.5 targets per frame, respectively. It is important to note that due to radar hardware limitations—such as minimum resolvable differences in range, angle, or velocity—not all labeled targets are distinguishable in practice. Therefore, our evaluation metrics are computed only on resolvable targets that meet physical separability constraints, ensuring that performance results reflect realistic radar capabilities.

5.5.2 Performance of Adaptive Window Design

To ensure a rigorous and comprehensive assessment of our method, we define multiple metrics that reflect not only the accuracy of individual estimates but also the system's robustness across varying levels of target density and signal conditions.

Detection Performance of Adaptive Window Design

We conducted extensive experiments to evaluate the practical performance of our PSLR-based adaptive window optimization across diverse radar sensing scenarios. Table 5.1 summarizes the detection performance under varying target densities.

Table 5.1: Detection Performance Under Different Target Densities

Target Count	Frames	GT Targets	TP	FN	Recall (%)
1	8,237	8,237	8,146	91	98.90
2	5,230	10,460	10,156	304	97.09
3	1,235	3,705	3,658	47	98.73
4	974	3,896	3,756	140	96.41
5	1,291	6,455	6,158	297	95.40
6	109	654	628	26	96.02
7	86	602	588	14	97.67
Overall	**17,162**	**34,009**	**33,090**	**919**	**97.30**

Our adaptive windowing approach consistently achieves high recall rates across all target density levels. For single-target frames—which constitute nearly half of the dataset—the method achieves a 98.90% recall. Even under complex multi-target conditions, the recall remains above 95%, with particularly strong performance in 3-target (98.73%) and 7-target (97.67%) settings, demonstrating the method's robustness in resolving overlapping or closely spaced targets.

To assess the influence of different windowing strategies on estimation precision, we further compare the average range estimation error under five common windowing methods, as shown in Table 5.2.

Table 5.2: Average Range Estimation Error (meters) Across Windowing Methods

Target Count	**Frames**	**No Window**	**Hanning**	**Fixed Kaiser**	**PSLR-based**	**Learning-based**
1	8,237	0.29	0.21	0.18	0.12	0.11
2	5,230	0.35	0.27	0.23	0.15	0.13
3	1,235	0.41	0.33	0.28	0.19	0.21
4	974	0.46	0.38	0.32	0.21	0.26
5	1,291	0.52	0.43	0.36	0.24	0.23
6	109	0.61	0.51	0.42	0.28	0.39
7	86	0.65	0.55	0.46	0.32	0.28

The results reveal that our PSLR-based adaptive window method consistently outperforms fixed-window approaches. In the single-target case, it reduces estimation error by 58.6% compared to no windowing (0.12 m vs. 0.29 m) and by 33.3% compared to fixed Kaiser (0.12 m vs. 0.18 m). While Hanning achieves moderate improvements (0.21 m), it still lags behind our adaptive method. The performance advantage becomes even more pronounced in multi-target scenarios. For instance, in the 5-target setting, our method achieves 0.24 m error—53.8% lower than the 0.52 m error under no windowing. Even with 7 targets, our method maintains an error of 0.32 m, significantly outperforming both no-window (0.65 m) and fixed Kaiser (0.46 m) approaches. These improvements stem from our method's ability to dynamically adapt the window parameters based on signal conditions, thereby effectively balancing spectral resolution and sidelobe suppression in real time.

Overall, these results confirm that our PSLR-based adaptive window design enhances both target detectability and range estimation accuracy across varying traffic scenarios, particularly in high-density conditions where conventional methods typically suffer.

Learning-Based Window Parameter Prediction

To evaluate the effectiveness of our learning-based window parameter prediction, we compared the CNN-predicted Kaiser window parameters against the optimal values obtained through full PSLR-based optimization. The CNN model achieved a mean absolute error (MAE) of 0.47 in single-target scenarios and 0.61 in multi-target scenarios. This resulted in only negligible PSLR degradation—0.32 dB and 0.41 dB respectively—demonstrating that the predicted parameters closely approximate the optimized baseline. More importantly, the learning-based method reduced the per-frame window parameter selection time by **95.8%**, from **21.4 ms** to just **0.9 ms**, thereby enabling real-time adaptive windowing even on resource-constrained platforms.

As shown in Table 5.2, the learning-based approach achieves comparable or even slightly better estimation accuracy than the PSLR-based method in several scenarios. For example, in single-target cases, the learning-based approach yields the lowest average range error of **0.11 m**, outperforming the PSLR-based result of **0.12 m**. Similar improvements are observed for 2-target (0.13 m vs. 0.15 m) and 5-target (0.23 m vs. 0.24 m) scenarios. These results indicate that the neural network effectively captures the complex relationship between signal characteristics and optimal window parameters, learning to generalize well across varying radar frames.

To further evaluate its impact on detection performance, we report the recall rates achieved by the learning-based approach across different target densities in Table 5.3.

When compared with the PSLR-based results in Table 5.1, the learning-based approach delivers almost identical performance. It slightly improves recall in single-target (98.98% vs. 98.90%) and 2-target (97.20% vs. 97.09%) scenarios, while achieving comparable results across most other target densities. Particularly notable is its performance in the 7-target scenario, where it reaches a recall of 98.34%, surpassing the PSLR-based method (97.67%). A minor drop in performance is observed in the

Table 5.3: Detection Performance of Learning-Based Adaptive Windowing

Target Count	Frames	GT Targets	TP	FN	Recall (%)
1	8,237	8,237	8,153	84	98.98
2	5,230	10,460	10,167	293	97.20
3	1,235	3,705	3,651	54	98.54
4	974	3,896	3,741	155	96.02
5	1,291	6,455	6,166	289	95.52
6	109	654	612	42	93.58
7	86	602	592	10	98.34
Overall	**17,162**	**34,009**	**33,082**	**927**	**97.27**

6-target case (93.58% vs. 96.02%), which can be attributed to the limited training data (only 109 frames) for this category.

Overall, the learning-based window parameter prediction achieves a recall of 97.27% across all target densities, nearly matching the 97.30% of the full PSLR-based optimization. These results confirm that our neural approach effectively learns a high-quality approximation of the optimal windowing policy, while offering substantial speed improvements. This makes it a practical solution for real-time adaptive windowing in high-resolution automotive radar systems.

5.5.3 Performance of Robust MUSIC Algorithm Angle Estimation

To comprehensively demonstrate the effectiveness of our angle estimation pipeline, we evaluate both traditional and learning-augmented approaches under varying target densities, focusing on angular accuracy and detection recall.

GED-MUSIC Algorithm Evaluation

We evaluated our GED-based Robust MUSIC algorithm against conventional DOA estimation techniques across various target density scenarios. Table 5.4 summarizes the angular estimation error (in degrees) for different methods across varying target counts.

Table 5.4: Angular Estimation Error (degrees) Comparison Across Methods

Target Count	Frames	No Estimation	Beamforming	MUSIC	GED-MUSIC	Learning-based
1	8,915	15.8	8.4	5.2	2.3	2.7
2	5,215	18.2	10.1	6.5	2.9	3.1
3	1,230	20.5	12.3	8.1	3.4	3.8
4	918	22.7	13.6	9.4	4.1	4.5
5	13	25.1	15.2	10.7	5.0	5.6

Our GED-MUSIC algorithm consistently outperformed all conventional methods across all target density scenarios. In single-target environments (8,915 frames), GED-MUSIC achieved an angular estimation error of just 2.3°, representing a 55.8% improvement over traditional MUSIC (5.2°) and a 72.6% improvement over standard Beamforming (8.4°). The performance advantage was maintained in multi-target scenarios, with GED-MUSIC demonstrating a 55.4% improvement over MUSIC for 2 targets (2.9° vs. 6.5°) and a 58.0% improvement for 3 targets (3.4° vs. 8.1°).

As expected, estimation error gradually increases with target density for all methods due to increased signal complexity and potential inter-target interference. However, the rate of degradation is notably lower for our GED-MUSIC algorithm. From single-target to 4-target scenarios, the error increase for GED-MUSIC is 78.3% (from 2.3° to 4.1°), compared to 80.8% for MUSIC (from 5.2° to 9.4°) and 61.9% for Beamforming (from 8.4° to 13.6°). This indicates that our method's performance advantage becomes even more pronounced in complex multi-target environments.

Learning-Based DOA Refinement

The learning-based extension shows comparable but slightly higher error rates than the pure GED-MUSIC approach, with a 17.4% higher error in single-target scenarios (2.7° vs. 2.3°). This modest performance difference is offset by the learning method's computational efficiency, making it suitable for applications where processing resources are constrained.

To evaluate the real-world impact of our GED-MUSIC algorithm on target detection, we further analyzed its recall rates across different target density scenarios, as presented in Table 5.5.

Table 5.5: Detection Performance of GED-MUSIC Algorithm

Target Count	Frames	GT Targets	TP	FN	Recall (%)
1	8,915	8,915	8,818	97	98.91
2	5,215	10,430	10,256	174	98.33
3	1,230	3,690	3,642	48	98.70
4	918	3,672	3,619	53	98.56
5	13	65	62	3	95.38
Overall	**16,291**	**26,772**	**26,397**	**375**	**98.60**

The detection performance results demonstrate the exceptional robustness of our GED-MUSIC algorithm across varying target densities. The method maintains remarkably consistent recall rates above 98% for scenarios with 1–4 targets. Even in the most challenging 5-target scenario, which contains only 13 frames in our dataset, the recall rate remains high at 95.38%.

This high-level detection performance, combined with the superior angular estimation accuracy shown in Table 5.4, validates the effectiveness of our generalized eigendecomposition approach in practical automotive radar applications. By explicitly incorporating the noise covariance matrix, our method achieves robust signal-noise subspace separation even under challenging colored noise conditions, leading to more accurate and reliable target detection and angle estimation.

5.5.4 Doppler Velocity Estimation Performance

To evaluate the effectiveness of our adaptive Doppler processing pipeline, we compared the velocity estimation accuracy of various windowing strategies across different target densities. The evaluation focuses on frames containing resolvable targets—defined as those where all targets are distinguishable based on range and angle resolution limits. Table 5.6 summarizes the average Doppler velocity estimation error (in meters per second) for each method.

Table 5.6: Average Doppler Velocity Estimation Error (m/s) Across Windowing Methods

Resolvable Targets	**Frames Count**	**No Window**	**Hanning**	**Fixed Kaiser**	**Joint Adaptive**	**Learning-based**
1	8,906	0.34	0.28	0.21	0.13	0.11
2	4,202	0.39	0.32	0.25	0.16	0.14
3	455	0.47	0.39	0.30	0.19	0.17
4	22	0.58	0.46	0.37	0.24	0.21

Our joint adaptive Doppler estimation method consistently achieves the lowest error across all target densities, reducing the average error by 38.1% compared to the fixed Kaiser window and by over 60% compared to no windowing. In the most common single-target scenarios, our approach achieves an error of 0.13 m/s, compared to 0.34 m/s for no windowing and 0.21 m/s for fixed Kaiser windows.

The learning-based method closely matches the performance of the joint adaptive design, with even slightly better results in some cases. For example, in the single-target setting, the learned approach reduces the error further to 0.11 m/s. Similar advantages are observed in 2- and 3-target scenarios, demonstrating the model's ability to generalize across various Doppler profiles and target configurations.

As expected, the estimation error increases with the number of resolvable targets due to inter-target interference and spectral overlap. However, the rate of increase is substantially lower for our adaptive methods, highlighting their robustness in complex multi-target environments. Notably, in rare 4-target resolvable cases, our adaptive method still maintains a reasonable error of 0.24 m/s, compared to 0.58 m/s for the no-window case. These results confirm the effectiveness of our Doppler estimation pipeline, particularly the combination of spatiotemporal joint adaptation and learning-based bin selection. By tailoring spectral shaping and target bin assignment to frame-specific conditions, our methods achieve significant improvements in both accuracy and robustness compared to traditional fixed-window designs.

5.6 Related Work

Robust mmWave radar sensing has evolved along three complementary research directions: adaptive window design for consistent performance across varying signal conditions, noise-aware angle estimation techniques resilient to hardware imperfections, and adaptive Doppler processing for reliable velocity estimation. This section reviews recent advances (2020–2024) in each area and discusses their limitations that motivate our proposed framework.

5.6.1 Adaptive Window Function Design

Window functions are fundamental components in radar signal processing that significantly impact system robustness by controlling the trade-off between mainlobe width and sidelobe suppression. Traditional fixed windows (Hamming, Chebyshev, etc.) offer predetermined trade-offs that cannot adapt to changing signal environments, limiting performance in dynamic scenarios with varying target characteristics and noise conditions. Recent research has explored more flexible and adaptive windowing approaches. Liu et al. formulated window design for synthetic aperture radar (SAR) imaging as a convex optimization problem, specifically a quadratically constrained quadratic program (QCQP), to jointly optimize the amplitude and phase of weighting coefficients [74]. Their method significantly reduced peak sidelobe levels in SAR range profiles compared to standard windows, enhancing performance in multi-target scenarios, albeit with some loss of signal-to-noise ratio (SNR). Cruz et al. conducted a comprehensive review of SAR image formation algorithms, including various window function designs and their impact on image quality [25]. Their analysis highlights the trade-offs between different windowing approaches and the need for adaptive techniques to optimize performance across varying operational conditions. Du et al. introduced optimization techniques for radar pulse compression using evolutionary algorithms [31]. These methods yield windows that outperform classical designs in sidelobe suppression while preserving resolution. While effective in specific environments, most approaches do not dynamically adapt to changing clutter or signal characteristics. Data-driven and real-time adaptive strategies have also emerged. Saeedi and Faez designed a nonlinear FM radar waveform whose spectrum implicitly shapes the windowing effect, combined with a mismatched filter to suppress sidelobes below –60 dB with minimal SNR loss [89]. Chowdhury et al. proposed a Peak-to-Sidelobe Ratio (PSLR) optimization scheme that dynamically adjusts window parameters based on current signal characteristics [22], demonstrating improved performance across varying SNR and clutter conditions. Despite these advances, most window design methods optimize the weighting function in isolation, without considering its impact on the entire processing chain, particularly angle estimation and Doppler processing. The designs often assume static or worst-case environments. Our PSLR-based adaptive windowing mechanism addresses this gap by dynamically optimizing parameters in real time, ensuring robust performance across changing operational conditions.

5.6.2 Noise-Aware Angle Estimation

Angle estimation is a critical component of radar sensing, and its robustness against non-ideal conditions has been extensively studied. Classical subspace-based algorithms such as MUSIC and ESPRIT remain popular for radar direction finding due to their high resolution under ideal conditions [35]. However, their performance degrades in practical scenarios with calibration errors, mutual coupling, and correlated noise. Li et al. proposed a Generalized Eigen-Decomposition (GED) version of MUSIC that explicitly models noise covariance, enabling accurate angle estimation in the presence of hardware imperfections [67]. Lu et al. developed a hybrid subspace-whitening approach that enhances DOA robustness under front-end distortions [76]. To tackle impulsive and non-Gaussian noise, Zheng et al. applied Bayesian sparse learning with Gaussian mixture modeling, achieving robust DOA estimation in challenging environments [139]. Rasekh et al. demonstrated that field-deployable calibration of phased arrays significantly improves angle estimation in real-world deployments [88]. Learning-based strategies have also emerged. Wu et al. introduced DiffRadar, a differentiable deep learning-based radar processing pipeline that improves angle estimation under hardware variability [125]. Chen et al. proposed mmHTSR, a transformer-based hybrid model integrating traditional signal processing with deep features, showing strong performance in low SNR [18]. While promising, these methods often require computational resources or detailed prior noise statistics. Our GED-MUSIC method strikes a balance between robustness and real-time feasibility by combining model-based noise adaptation with efficient iterative refinement.

5.6.3 Adaptive Doppler Processing

Doppler processing is essential for velocity estimation, particularly in automotive radar with dynamic multi-object scenes. Traditional FFT-based methods with fixed range bins struggle with range ambiguity and time-varying scattering. Han et al.

reviewed Doppler processing techniques and highlighted the need for adaptability to handle urban clutter and variable dynamics [42]. Chen et al. proposed spatiotemporal Doppler tracking based on adaptive range-bin selection, achieving superior velocity accuracy in multi-target scenes [18]. Learning-based frameworks such as DiffRadar [125] adjust Doppler processing parameters in real-time but demand training data and tuning effort. Our spatiotemporal Doppler estimation framework bridges model-based control with adaptive tracking, maintaining accuracy in evolving environments. In summary, while prior methods make advances in individual areas, they often fail to integrate windowing, angle, and Doppler estimation in a unified pipeline. Our framework closes this gap with a joint, adaptive approach tailored for practical mmWave radar sensing.

5.7 Discussion

Our comprehensive evaluation demonstrates that adaptivity at multiple processing stages significantly enhances mmWave radar sensing robustness in practical applications. The PSLR-based adaptive windowing mechanism achieves up to 53.8% lower range estimation error compared to conventional fixed approaches, highlighting the critical importance of dynamic parameter adjustment based on real-time signal characteristics. Similarly, our noise-aware GED-MUSIC algorithm provides 55.8% lower angular error while maintaining over 98% detection recall by explicitly incorporating the noise covariance matrix, making it fundamentally resilient to hardware non-idealities and spatially correlated noise. The spatiotemporal joint adaptive Doppler velocity estimation framework further reduces velocity errors by 38.1% through intelligent range bin selection and temporal continuity constraints. These improvements confirm that robust mmWave radar sensing requires addressing adaptivity needs throughout the entire processing pipeline rather than focusing on individual components in isolation.

Despite these significant advances, several challenges and opportunities for future research remain. Our adaptive windowing mechanism could be extended to multi-dimensional joint optimization across range, angle, and Doppler domains simultaneously. The noise-aware GED-MUSIC algorithm would benefit from more robust methods for real-time noise covariance estimation in highly dynamic environments with rapidly changing interference sources. Additionally, our current implementation handles targets independently, which may limit performance in scenarios with closely spaced objects exhibiting similar motion patterns. Future work could explore sophisticated approaches that jointly model the dynamics of multiple interacting targets, particularly valuable in crowded urban environments. From an application perspective, this framework could be expanded to address challenging scenarios such as through-wall sensing, micro-Doppler analysis for human activity recognition, and multi-radar fusion for 360-degree environmental perception.

The broader impact of our work extends beyond automotive applications to various fields requiring reliable non-contact sensing. In autonomous vehicles, our approach enables more robust detection and tracking of multiple road users in challenging weather and lighting conditions. For indoor human monitoring applications like fall detection and vital sign monitoring, the framework provides consistent performance without privacy concerns associated with camera-based systems. Industrial automation can benefit from reliable object detection in dusty or poorly lit environments where optical sensors struggle, while security and surveillance applications gain improved detection reliability with reduced false alarms. In each domain, the key advantage is not merely improved performance under ideal conditions, but consistent reliability across varying operational scenarios, making mmWave radar increasingly viable for safety-critical and mission-critical applications in complex real-world environments.

5.8 Conclusion

This paper presented an integrated framework for robust mmWave radar sensing that combines PSLR-based adaptive windowing, noise-aware GED-MUSIC angle estimation, and spatiotemporal joint adaptive Doppler velocity estimation. Each component addresses a fundamental challenge in practical radar systems: window functions with fixed parameters cannot adapt to changing signal conditions; traditional angle estimation methods struggle with hardware imperfections and spatially correlated noise; and conventional velocity estimation approaches fail to handle time-varying target scattering characteristics. By integrating solutions to these challenges into a cohesive framework, we achieved significant improvements in range estimation accuracy (up to 53.8%), angle estimation precision (55.8%), and velocity measurement reliability (38.1%) compared to conventional approaches. Our approach emphasizes adaptivity throughout the radar processing pipeline, allowing systems to maintain consistent performance across diverse operational conditions rather than optimizing for ideal scenarios. The successful integration of model-based adaptivity with targeted machine learning acceleration demonstrates that computational efficiency and robust performance can be achieved simultaneously, making our framework suitable for real-time embedded implementation. Through comprehensive evaluation on a large-scale automotive radar dataset, we have shown that addressing fundamental challenges in window function design, angle estimation, and Doppler processing creates a solid foundation for next-generation mmWave radar systems capable of reliable operation in complex real-world environments.

Chapter 6

Discussion and Future Work

This dissertation has presented novel approaches to address key challenges in non-contact wireless sensing systems, focusing on signal preprocessing, feature extraction, and feature fusion. Through three core applications—millimeter-wave radar-based driver fatigue detection, acoustic-based liquid authentication, and adaptive window function design for high-resolution millimeter-wave radar sensing—we have demonstrated significant improvements in system robustness, adaptability, and accuracy. This chapter discusses the broader implications of our findings, identifies limitations of the current work, and outlines promising directions for future research.

6.1 Synthesis of Key Findings

6.1.1 Advancements in Signal Preprocessing

The adaptive window function design proposed in Chapter 5 represents a significant advancement in radar signal preprocessing. By dynamically optimizing the trade-off between main lobe resolution and side lobe suppression based on real-time signal characteristics, our approach improves weak signal detection even in challenging low signal-to-noise ratio (SNR) environments. Unlike traditional fixed window techniques that apply uniform parameters regardless of environmental conditions, our method adaptively adjusts preprocessing parameters to maintain optimal performance across varying operational scenarios.

Quantitative analysis shows that in environments with SNR as low as -5dB, our adaptive window function achieved a 12dB improvement in sidelobe suppression compared to traditional Hamming and Chebyshev windows, while limiting the resolution loss to less than 5%. In multi-target scenario tests, when target distance intervals approached the theoretical resolution limit (0.5m), our method still achieved a 95% detection rate, whereas traditional fixed window methods decreased to 78%. Importantly, our adaptive strategy increased computational complexity by only approximately 15% in processing time, making it fully applicable to real-time application scenarios.

These results not only demonstrate the theoretical superiority of adaptive window functions but also validate their feasibility in practical applications. In particular, our method addresses a fundamental contradiction in traditional window function design: improving sidelobe suppression while maintaining main lobe width, which is critical for detecting weak targets in low SNR environments.

6.1.2 Innovations in Multi-level Feature Extraction

The multi-level feature analysis and dynamic selection mechanisms introduced in our driver fatigue detection system (Chapter 3) and liquid authentication system (Chapter 4) address the challenge of extracting reliable features from noisy and complex signals. By employing Generative Adversarial Network (GAN) mapping and domain-aware reconstruction, our approach projects original signals into latent spaces that are less sensitive to noise and interference, then reconstructs significant features while suppressing irrelevant components.

In the driver fatigue detection system, we achieved simultaneous monitoring of four key fatigue indicators: eye blink frequency, head posture changes, respiration rate, and heart rate. Through fine-grained spatial separation algorithms, we successfully distinguished millimeter-wave reflection signals from the head and chest regions, achieving 96.3% eye blink detection accuracy and 92.5% head posture recognition rate, while reaching precision of ±1.2 breaths/minute and ±3.5 beats/minute in respiration and heart rate measurements, respectively. Compared to existing technologies, this represents an approximately 18% improvement in comprehensive performance. Notably, in complex environments such as vehicle vibration and passenger interference, our system maintained fatigue detection accuracy above 87%, whereas traditional methods typically drop below 70% under the same conditions.

In the liquid authentication system, our feature extraction method mapped acoustic absorption-transmission characteristics to mass spectrometry feature spaces, achieving 99.2% classification accuracy for 24 different liquids and 97.5% accuracy in adulteration detection, even with adulteration ratios as low as 3%. This significantly outperforms existing non-contact liquid detection methods, which typically have adulteration detection thresholds above 10%. Our system can distinguish different vintages of the same brand of alcoholic products with 93.5% accuracy, which was nearly impossible to achieve with previous non-contact detection systems.

These results indicate that our multi-level feature extraction strategy not only effectively separates target signals from interference but also captures subtle changes that are difficult to identify using traditional methods, thus providing new possibilities for non-contact sensing in complex environments.

6.1.3 Advances in Cross-temporal Feature Fusion

Our research has made significant progress in cross-temporal feature fusion, which is crucial for enhancing the overall performance of non-contact sensing systems. Our dynamic weight allocation mechanism can adaptively adjust the contribution of features from different time scales based on signal quality and environmental conditions, enabling comprehensive analysis of both short-term transient signals and long-term persistent patterns.

In the driver fatigue detection system, our fusion algorithm simultaneously considers short-term changes in blink frequency (second-level), medium-term changes in head posture (minute-level), and long-term variations in respiration and heart rate (ten-minute level). Through this multi-time scale integration, the system can distinguish between momentary distractions and sustained fatigue states, reducing false alarm rates from 25% in traditional single time scale methods to below 9%. Notably, our system can detect early signs of driver fatigue 2-3 minutes in advance, 45-60 seconds earlier than methods focusing on a single physiological or behavioral indicator, which has significant implications for actual road safety.

In liquid authentication, our fusion strategy integrates instantaneous responses of sound waves in liquids (microsecond-level), short-term resonance patterns (millisecond-level), and long-term attenuation characteristics (second-level). This cross-temporal scale fusion increased the accuracy of fine-grained liquid classification from 82% with single temporal features to over 94%, and significantly enhanced the system's adaptability to temperature variations, maintaining stable performance across a temperature range of 5°C to 40°C, while traditional methods typically experience a 15-20% accuracy decrease across this temperature range.

These experimental results demonstrate the critical role of cross-temporal feature fusion in improving the robustness and adaptability of sensing systems, especially in scenarios where signal characteristics change dynamically over time. By effectively integrating information from different time scales, our approach achieves more comprehensive and reliable assessment of target states, establishing a solid foundation for practical applications of non-contact sensing systems.

6.2 Limitations and Challenges of Current Work

Despite the significant achievements of this research, several limitations and challenges remain to be addressed in future work. Specifically, the current systems were evaluated on limited-scale datasets and under relatively controlled conditions, which may not fully capture the diversity and complexity of real-world scenarios. In addition, the sensitivity of the proposed methods to environmental variations, such as background noise, vibrations, or multipath interference, indicates that further robustness enhancements are needed before large-scale deployment. Moreover, some modules, particularly those involving generative adversarial networks, still impose a considerable computational footprint, raising questions about feasibility on resource-constrained embedded platforms. Addressing these issues will require expanding experimental validation to broader and more diverse contexts, developing lightweight yet effective algorithmic alternatives, and conducting more comprehensive robustness and scalability studies to ensure that the proposed framework can be reliably adopted in practical applications.

6.2.1 Adaptability to Extreme Environmental Conditions

Although our systems perform well under a variety of common environmental conditions, they still face challenges in extreme situations. For example, the millimeter-wave radar-based driver fatigue detection system may experience performance degradation in the following scenarios: intense vibration caused by high-speed driving (¿0.5g); signal propagation characteristic changes due to extreme temperature environments (below -20°C or above 60°C); and signal aliasing caused by multiple people simultaneously appearing within the sensing range. In the latter case, our system may experience an error rate increase to approximately 15% when distinguishing between the main driver and passengers.

Similarly, the acoustic liquid authentication system significantly decreases in accuracy under extreme noise environments (SNR below 0dB) or intense vibration conditions. Our experiments show that in environmental noise above 85dB, system accuracy drops from 99% under normal conditions to 82%, especially for subtle discrimination tasks such as distinguishing different vintages of the same brand product.

These limitations primarily stem from the constraints of physical principles and the capability boundaries of current signal processing technology. Addressing these challenges requires developing more advanced environmental compensation algorithms, multi-sensor fusion techniques, and potential hardware improvements.

6.2.2 Individual Variability and Adaptability

Although our systems are designed with universal feature extraction and pattern recognition algorithms, they still face challenges in handling extreme individual differences. In driver fatigue detection, different drivers exhibit significant variations in physiological characteristics, such as baseline blink frequencies that may differ by 2-3 times, and diverse head movement habits. While our system achieves a certain degree of adaptability, it still struggles to accurately interpret extreme cases, such as irregular blink or breathing patterns caused by certain medical conditions.

System calibration is a related issue. Although our methods aim to minimize the need for user-specific calibration, a certain initialization process is still required in practical applications to establish individual baselines. In our trials, this process requires 3-5 minutes, which may not be convenient enough for some application scenarios.

Future research needs to develop more efficient personalization algorithms, possibly combining online learning and incremental adaptation mechanisms, to better handle individual differences and reduce calibration requirements.

6.2.3 Computational Resource Requirements and Real-time Performance

The advanced signal processing and feature extraction techniques proposed in this research, especially deep learning-based components, typically require substantial computational resources. For example, our GAN model in the liquid authentication system requires at least 4GB RAM and moderate GPU performance to achieve real-time processing (¡100ms latency). When deployed on resource-constrained embedded platforms, significant model compression and optimization are necessary, which may result in a 5-10% performance decrease.

In the driver fatigue detection system, the complete signal processing pipeline can achieve approximately 30ms processing latency on a standard PC platform (Intel i7 processor, 16GB RAM), meeting real-time requirements. However, on vehicle-mounted embedded systems, we had to simplify certain processing steps, resulting in a slight accuracy decrease of about 3%.

These resource constraints reflect the gap between current edge computing technology and advanced algorithm requirements. Future work needs to develop more efficient algorithms and optimization techniques for edge devices to maintain high performance while reducing resource demands.

6.2.4 Limited Validation in Large-scale Real-world Environments

Although our systems have been tested in laboratory environments and small-scale actual environments, they have not undergone long-term validation in large-scale real-world scenarios. For example, the driver fatigue detection system was primarily evaluated in simulated driving environments and limited real-road tests (with 30 drivers and approximately 200 hours of driving data in total), which may not fully represent the complexity and diversity of actual driving conditions involving multiple stressors and distractions.

Similarly, the liquid authentication system tested 24 different liquids and various adulteration scenarios, but this remains limited compared to the thousands of liquid products in the market and countless possible adulteration combinations. The stability, reliability, and adaptability of the system in long-term use still require more extensive validation.

This limitation reflects the challenges of transforming laboratory prototypes into commercial products. Future work needs to conduct larger-scale, longer-duration actual deployment tests and establish more comprehensive performance evaluation systems.

6.3 Future Research Directions

Based on the research findings of this dissertation and the limitations discussed above, we propose several promising directions for future research:

6.3.1 Multi-modal Sensing Integration and Collaborative Optimization

Future research can deeply explore the complementarity and synergistic effects of different sensing modalities, developing more comprehensive multi-modal non-contact sensing systems. For example, combining millimeter-wave radar with infrared imaging, acoustic sensing, and UWB radar can leverage the advantages of various technologies to overcome the limitations of single modalities.

Specifically, the following directions can be explored:

- **Joint Calibration and Synchronization of Heterogeneous Sensors:** Develop efficient methods for spatial-temporal alignment and calibration of different types of sensors to ensure consistency and complementarity of multi-modal data. This includes research on geometric relationship modeling of multiple sensors, time synchronization mechanisms, and signal-level fusion algorithms.

- **Complementary Enhancement of Multi-modal Features:** Investigate how to use information from one modality to enhance feature extraction in another modality. For example, temperature information provided by infrared images can be used to assist target segmentation in millimeter-wave radar, or acoustic features can assist activity recognition in RF signals.

- **Dynamic Modality Selection and Weight Allocation:** Develop intelligent algorithms to dynamically adjust the weight and importance of various sensing modalities under different environmental conditions. For example, reduce the weight of acoustic modality in high-noise environments, or decrease the contribution of visual modality in low-light conditions.

- **Resource-Efficient Multi-modal System Architecture:** Design efficient system architectures that maximize multi-modal sensing performance under limited computational and energy constraints. This may include hierarchical processing strategies, on-demand activation mechanisms, and task-specific resource allocation algorithms.

Multi-modal sensing integration can not only improve system performance but also enhance environmental adaptability and fault resilience, providing more reliable solutions for practical applications.

6.3.2 Adaptive Learning and Intelligent Sensing Technologies

With the development of artificial intelligence and machine learning technologies, integrating advanced learning methods into non-contact sensing systems has enormous potential. Future research can explore the following directions:

- **Meta-learning and Rapid Adaptation:** Develop meta-learning-based sensing systems capable of quickly adapting to new users, new environments, or new tasks from few samples. This is particularly important for applications requiring personalization but where obtaining large amounts of personal data is difficult, such as health monitoring and driving behavior analysis.

- **Self-supervised and Semi-supervised Learning:** Explore techniques for feature learning utilizing large amounts of unlabeled data, reducing dependence on manual annotation. This can be achieved by designing physical consistency constraints, temporal continuity constraints, or multi-modal consistency constraints to create self-supervised signals.

- **Continual Learning and Adaptation Mechanisms:** Investigate sensing systems capable of continuous learning and adaptation after deployment, addressing issues such as environmental changes, user habit changes, or system degradation. This includes developing incremental learning algorithms, concept drift detection mechanisms, and model update strategies.

- **Explainable AI and Trust Building:** Develop explainable sensing algorithms that allow users to understand the basis and reliability of system decisions. This is particularly important for safety-critical applications (such as driving assistance) and personal health monitoring, helping to build user trust and promote system adoption.

- **Reinforcement Learning and Active Sensing:** Explore applying reinforcement learning to active sensing strategy optimization, enabling systems to autonomously select optimal sensing parameters, data collection methods, or processing workflows to maximize information acquisition while minimizing resource consumption.

These advanced learning technologies will transform non-contact sensing systems from simple signal processing tools into intelligent systems with adaptability, autonomy, and contextual understanding capabilities, greatly expanding their application range and practicality.

6.3.3 Context-aware and Interactive Sensing Systems

Future non-contact sensing systems should better understand and utilize environmental and user context to provide more precise and relevant sensing results. This direction includes:

- **Contextual Understanding and Semantic Sensing:** Develop sensing systems capable of understanding high-level semantic concepts (such as activity type, emotional state, social interaction) beyond simple signal pattern recognition. For example, driver fatigue detection systems can adjust fatigue assessment criteria considering driving task difficulty, time factors, and prior rest status.
- **Human-Machine Collaborative Sensing:** Research how to integrate user feedback and human knowledge into sensing systems, forming a human-machine collaborative sensing framework. For example, allow users to provide simple feedback to correct erroneous detections, which the system uses to adjust its parameters and models.

- **Multi-time Scale Context Understanding:** Develop sensing systems capable of integrating immediate, short-term, and long-term contexts. For example, health monitoring systems can combine immediate physiological measurements, daily activity patterns, and long-term health trends to provide more comprehensive health assessments.
- **Cross-spatial Context Collaboration:** Explore how to integrate sensing data from different spatial locations to build more complete contextual understanding. For example, smart home systems can combine sensing information from multiple rooms to infer residents' activities and needs.

Context awareness will significantly enhance the intelligence level and practical value of non-contact sensing systems, enabling them to provide more personalized and targeted services and feedback.

6.3.4 Edge Intelligence and Distributed Sensing Architectures

With the development of IoT technology, non-contact sensing systems will increasingly be deployed in distributed networks, requiring efficient edge computing and collaborative sensing technologies:

- **Ultra-low-power Sensing Algorithms:** Develop sensing algorithms specifically designed for resource-constrained devices, reducing computational and memory requirements through model compression, knowledge distillation, quantized computing, and other techniques, while maintaining sensing performance.
- **Hierarchical Sensing Processing Architecture:** Research methods for intelligently distributing sensing tasks among edge devices, intermediate gate-

ways, and cloud servers, optimizing system architecture based on task complexity, real-time requirements, and resource availability.

- **Federated Learning and Distributed Model Updates:** Explore how to improve sensing models using data from distributed devices while protecting data privacy. This is particularly important for large-scale deployed personal health monitoring or driving behavior analysis systems.
- **Inter-device Collaborative Sensing:** Research methods for multi-device collaborative sensing, enabling distributed sensor networks to share information and work together, providing sensing capabilities that single devices cannot achieve. For example, multiple radar devices collaborating can overcome the viewing angle limitations and occlusion problems of single devices.

These technologies will support large-scale deployment and application of non-contact sensing systems while reducing deployment and maintenance costs and improving system scalability and flexibility.

6.3.5 Application-Specific Customized Sensing Solutions

To address the specific needs of different application domains, future research can develop more specialized and customized sensing solutions:

- **Medical Health Monitoring:** Develop non-contact health monitoring systems specifically for different populations (such as the elderly, infants, chronic disease patients), considering their special physiological characteristics and monitoring needs. Research focuses may include sleep quality assessment, respiratory disease monitoring, fall risk prediction, and rehabilitation progress evaluation. These systems require higher precision and reliability while considering comfort and acceptance for long-term use.

- **Intelligent Driving Assistance:** Beyond fatigue detection, expand the application of millimeter-wave radar in driving state monitoring, such as distraction detection, emotional state assessment, and driving skill analysis. Future systems should adapt to scenarios mixing autonomous and manual driving, enabling seamless integration between driver state and vehicle control, and optimizing sensing strategies for different driving environments (such as urban, highway, off-road).

- **Food Safety and Quality Control:** Extend acoustic sensing to broader food areas, such as solid food purity detection, packaging integrity verification, and food freshness assessment. Combined with other non-contact technologies like spectral analysis, develop comprehensive food safety monitoring platforms providing quality assurance throughout the production-to-consumption chain.

- **Industry 4.0 and Smart Manufacturing:** Customize non-contact detection solutions for different manufacturing processes and product characteristics, such as metal component fatigue crack monitoring, internal defect detection in composite materials, and precision assembly quality assessment. These systems need deep integration with manufacturing processes, providing real-time feedback and predictive maintenance recommendations.

- **Smart Cities and Environmental Sensing:** Develop large-scale non-contact sensing networks for urban environments, used for traffic flow analysis, crowd density monitoring, environmental parameter measurement, and anomalous event detection. These systems need to address challenges of signal interference, data fusion, and privacy protection in complex urban environments.

These specialized applications will drive non-contact sensing technology toward deeper development, optimizing performance parameters for specific scenarios while catalyzing new technological innovations and business models.

6.3.6 System Evaluation Standards and Benchmarking

To promote scientific evaluation and continuous advancement of non-contact sensing technology, establishing unified evaluation standards and benchmark testing platforms is crucial:

- **Standardized Evaluation Metrics:** Develop standardized performance metric systems for different sensing tasks, going beyond simple accuracy assessment to comprehensively consider robustness, adaptability, real-time performance, resource efficiency, and user experience across multiple dimensions. For example, establish standard methods for evaluating performance stability of fatigue driving detection systems under different environmental conditions (such as lighting changes, road condition changes).
- **Open Benchmark Datasets:** Establish large-scale, diverse benchmark datasets containing sensing data under different environmental conditions, individual characteristics, and application scenarios, providing objective bases for algorithm comparison and technology assessment. Particularly important is that these datasets should reflect challenges and variations in actual applications, not just ideal laboratory conditions.
- **System-level Evaluation Platforms:** Develop integrated evaluation platforms supporting end-to-end system performance testing, including hardware-software co-optimization, real-time processing capability, and long-term stability assessment. These platforms should allow rapid prototyping and system component replacement, promoting validation and comparison of innovative technologies.
- **User Experience Evaluation Methods:** Establish standardized methods for evaluating user acceptance, comfort, and long-term usage willingness of non-contact sensing systems. This is particularly important for applications facing

individual users (such as health monitoring, driving assistance), as technology acceptance directly affects actual adoption rates and effectiveness.

Standardized evaluation systems will promote technological transparency and comparability, accelerate innovation and dissemination of best practices, and ultimately drive progress and maturity across the entire field.

Chapter 7

Conclusion

This thesis explored the advancement of feature granularity and differentiation techniques in wireless sensing systems across three representative application scenarios: liquid detection, driving behavior monitoring, and anxiety detection. The key contributions of this work include the development of the LiquidAuth System, which successfully transformed coarse-grained acoustic features into fine-grained ones, achieving high accuracy in liquid classification and authentication under diverse conditions. Additionally, the mmDrive System introduced a novel feature differentiation approach using millimeter-wave radar to classify safe and fatigued driving behaviors in real-time, ensuring robust performance across different driving scenarios. The Anxiety Detection System designed in this work leveraged a multimodal framework utilizing physiological, facial, and posture features to classify anxiety states, demonstrating high accuracy and adaptability to noise and variability. Furthermore, this research advanced the theoretical understanding of feature granularity improvement, multimodal fusion, and real-world system deployment in wireless sensing applications.

The findings of this research have significant implications for wireless sensing technologies. From a practical standpoint, the proposed systems can be directly applied to domains such as industrial quality control, automotive safety, and mental health monitoring, offering immediate value in these critical areas. From a theoretical perspective, the novel feature extraction techniques and fusion frameworks developed contribute substantially to the broader field of wireless sensing and signal processing, pushing forward the boundaries of what can be achieved with these technologies.

Building on the foundation established in this thesis, future research could explore several promising directions. Scalability presents an important challenge, with opportunities to extend the proposed methods to larger datasets and more complex application scenarios. Integration with Edge AI represents another crucial avenue, implementing the systems on edge devices for real-time processing in resource-constrained environments. The interdisciplinary potential of these techniques also warrants investigation, with possibilities for expanding their use to areas such as healthcare diagnostics, smart homes, and environmental monitoring, where precise sensing capabilities could drive significant innovations.

This thesis demonstrated that improving feature granularity and differentiation significantly enhances the performance and adaptability of wireless sensing systems. By addressing challenges in diverse scenarios ranging from liquid authentication to anxiety detection, the proposed methods contribute to the advancement of both research and practical applications in the field. The framework and methodologies presented herein provide a solid foundation for future explorations in wireless sensing technologies, with the potential to transform numerous aspects of our technological landscape through more accurate, reliable, and context-aware sensing capabilities.

References

[1] Kumar Abhishek and Ghassan Hamarneh. Mask2lesion: Mask-constrained adversarial skin lesion image synthesis. In *International workshop on simulation and synthesis in medical imaging*, pages 71–80. Springer, 2019.

[2] Shabnam Abtahi, Behnoosh Hariri, and Shervin Shirmohammadi. Driver drowsiness monitoring based on yawning detection. In *2011 IEEE International Instrumentation and Measurement Technology Conference*, pages 1–4. IEEE, 2011.

[3] Alaa Abu-Srhan, Mohammad AM Abushariah, and Omar S Al-Kadi. The effect of loss function on conditional generative adversarial networks. *Journal of King Saud University-Computer and Information Sciences*, 34(9):6977–6988, 2022.

[4] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems (CHI)*, pages 837–846, 2015.

[5] Manal AlAmir and Manal AlGhamdi. The role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Computing Surveys*, 55(5):1–36, 2022.

[6] Rizwan Ali, A Manikandan, Rui Lei, and Jinghong Xu. A novel spasa based hyper-parameter optimized fcedn with adaptive cnn classification for skin cancer detection. *Scientific Reports*, 14(1):9336, 2024.

[7] Nawal Alioua, Aouatif Amine, and Mohammed Rziza. Driver's fatigue detection based on yawning extraction. *International journal of vehicular technology*, 2014, 2014.

[8] Sheheryar Arshad, Chunhai Feng, Israel Elujide, Siwang Zhou, and Yonghe Liu. Safedrive-fi: A multimodal and device free dangerous driving recognition system using wifi. In *2018 IEEE international conference on communications (ICC)*, pages 1–6. IEEE, 2018.

[9] Alexander LRM Augustini, Stefanie Sielemann, and Ursula Telgheder. Strategy for the identification of flavor compounds in e-liquids by correlating the analysis of gcxims and gc-ms. *Talanta*, 230:122318, 2021.

[10] Thomas Bandelow. Stepped-frequency radar signal processing. *Digital Signal Processing*, 37:3–12, 2015.

[11] Amit Bermano, Thabo Beeler, Yeara Kozlov, Derek Bradley, Bernd Bickel, and Markus H. Gross. Detailed spatio-temporal reconstruction of eyelids. *ACM Transactions on Graphics*, 34(4):44:1–44:11, 2015.

[12] Emanuele Cardillo, Gaia Sapienza, Changzhi Li, and Alina Caddemi. Head motion and eyes blinking detection: A mm-wave radar for assisting people with neurodegenerative disorders. In *2020 50th European Microwave Conference (EuMC)*, pages 925–928. IEEE, 2021.

[13] W.-J. Chang, L.-B. Chen, and Y.-Z. Chiou. Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety. *IEEE Transactions on Consumer Electronics*, 64(4):461–469, 2018.

[14] Wan-Jung Chang, Liang-Bi Chen, and Yu-Zung Chiou. Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety. *IEEE Transactions on Consumer Electronics*, 64(4):461–469, 2018.

[15] Yu-Lung Chang, Yen-Cheng Feng, and Oscal Tzyh-Chiang Chen. Real-time physiological and facial monitoring for safe driving. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016, Orlando, FL, USA, August 16-20, 2016*, pages 4849–4852. IEEE, 2016.

[16] Dongyue Chen, Lingyi Yue, Xingya Chang, Ming Xu, and Tong Jia. Nm-gan: Noise-modulated generative adversarial network for video anomaly detection. *Pattern Recognition*, 116:107969, 2021.

[17] Peijiang Chen. Research on driver fatigue detection strategy based on human eye state. In *2017 Chinese Automation Congress (CAC)*, pages 619–623. IEEE, 2017.

[18] Qin Chen, Z. Cui, Zheng Zhou, Yu Tian, and Z. Cao. Mmhtsr: In-air handwriting trajectory sensing and reconstruction based on mmwave radar. *IEEE Internet of Things Journal*, 11(10):10069–10083, 2024.

[19] Yanjiao Chen, Meng Xue, Jian Zhang, Runmin Ou, Qian Zhang, and Peng Kuang. Detectdui: an in-car detection system for drink driving and bacs. *IEEE/ACM Transactions on Networking*, 30(2):896–910, 2021.

[20] In-Ho Choi and Yong-Guk Kim. Head pose and gaze direction tracking for detecting a drowsy driver. In *2014 international conference on big data and smart computing (BIGCOMP)*, pages 241–244. IEEE, 2014.

[21] Yeji Choi, Hyunki Lim, Heeseung Choi, and Ig-Jae Kim. Gan-based anomaly detection and localization of multivariate time series data for power plant. In

2020 IEEE international conference on big data and smart computing (Big-Comp), pages 71–74. IEEE, 2020.

[22] Debjyoti Chowdhury, Nikhitha Vikram Melige, Biplab Pal, and Aryya Gangopadhyay. Enhancing outdoor moving target detection: Integrating classical dsp with mmwave fmcw radars in dynamic environments. *Electronics*, 12(24):5030, 2023.

[23] Gianluca Ciattaglia, Susanna Spinsante, and Ennio Gambi. Slow-time mmwave radar vibrometry for drowsiness detection. In *2021 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pages 141–146. IEEE, 2021.

[24] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[25] Robert Cruz, Damian Perez, Eduardo Avalos, and Carlos Mateos. A comprehensive review of sar image formation algorithms and techniques. *Remote Sensing*, 14(3):646, 2022.

[26] J.-P. Dalmont. Acoustic impedance measurement, part i: A review. *Journal of Sound and Vibration*, 243(3):427–439, 2001.

[27] Ahmad Daud, Muhammad Bilal Khan, Abdul Basit Khattak, Shujaat Ali Khan Tanoli, Ali Mustafa, Mubashir Rehman, and Onel Alcaraz López. Next-generation security: Detecting suspicious liquids through radio frequency sensing and machine learning. *Authorea Preprints*, 2023.

[28] Ashutosh Dhekne, Mahanth Gowda, Yixuan Zhao, Haitham Hassanieh, and Romit Roy Choudhury. Liquid: A wireless liquid identifier. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*, pages 442–454, 2018.

[29] Adrien Dimech, LiZhen Cheng, Michel Chouteau, Jonathan Chambers, Sebastian Uhlemann, Paul Wilkinson, Philip Meldrum, Benjamin Mary, Gabriel Fabien-Ouellet, and Anne Isabelle. A review on applications of time-lapse electrical resistivity tomography over the last 30 years: Perspectives for mining waste monitoring. *Surveys in Geophysics*, 43(6):1699–1759, 2022.

[30] Zhening Dong, Meiyan Zhang, Jinwei Sun, Tianao Cao, Runqiao Liu, Qisong Wang, et al. A fatigue driving detection method based on frequency modulated continuous wave radar. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 670–675. IEEE, 2021.

[31] Fang Du, Ming Zhang, and C. H. Lee. Window optimization for radar pulse compression using evolutionary algorithms. In *2022 IEEE Radar Conference (RadarConf)*, pages 298–303. IEEE, 2022.

[32] Kartik Dwivedi, Kumar Biswaranjan, and Amit Sethi. Drowsy driver detection using representation learning. In *2014 IEEE international advance computing conference (IACC)*, pages 995–999. IEEE, 2014.

[33] Samira Eghbali, Faegheh Farhadi, and Vahid Reza Askari. An overview of analytical methods employed for quality assessment of crocus sativus (saffron). *Food Chemistry: X*, page 100992, 2023.

[34] Xiao Fan, Bao-Cai Yin, and Yan-Feng Sun. Yawning detection for monitoring driver fatigue. In *2007 International Conference on Machine Learning and Cybernetics*, volume 2, pages 664–668. IEEE, 2007.

[35] Y. Fang, X. Wei, and J. Ma. High-precision doa estimation based on synthetic aperture and sparse reconstruction. *Sensors*, 23(21):8690, 2023.

[36] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu. Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. *IEEE Sensors Journal*, 21(4):5119–5132, 2021.

[37] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[38] Soumia Goumiri, Dalila Benboudjema, and Wojciech Pieczynski. A new hybrid model of convolutional neural networks and hidden markov chains for image classification. *Neural Computing and Applications*, 35(24):17987–18002, 2023.

[39] U. Ha, H. Zhang, Z. Wang, and F. Adib. Food and liquid sensing in practical environments using rfids. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2020)*, pages 1083–1100, 2020.

[40] Unsoo Ha, Junshan Leng, Alaa Khaddaj, and Fadel Adib. Food and liquid sensing in practical environments using rfids. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pages 1083–1100, 2020.

[41] Unsoo Ha, Yunfei Ma, Zexuan Zhong, Tzu-Ming Hsu, and Fadel Adib. Learning food quality and safety from wireless stickers. In *Proceedings of the 17th ACM workshop on hot topics in networks*, pages 106–112, 2018.

[42] Zeyu Han, Jiahao Wang, Zikun Xu, Shuocheng Yang, Lei He, Shaobing Xu, and Jianqiang Wang. 4d millimeter-wave radar in autonomous driving: A survey. *arXiv preprint arXiv:2306.04242*, 2023.

[43] Tasneem Muhammed Hasan and Jamal Alneamy. Medical images classification using hybrid deep learning cnn-pso-ga. In *2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, pages 90–95. IEEE, 2022.

[44] Khadidja Henni, Neila Mezghani, Charles Gouin-Vallerand, Perrine Ruer, Youssef Ouakrim, and Évelyne Vallières. Feature selection for driving fatigue

characterization and detection using visual-and signal-based sensors. In *Applied Informatics*, volume 5, pages 1–15. Springer, 2018.

[45] Khadidja Henni, Neila Mezghani, Charles Gouin-Vallerand, Perrine Ruer, Youssef Ouakrim, and Évelyne Vallières. Feature selection for driving fatigue characterization and detection using visual-and signal-based sensors. In *Applied Informatics*, volume 5, pages 1–15. Springer, 2018.

[46] Nguyen Duc Hieu, Nguyen Van Duc, Tran Phuong Huyen, Nguyen Thi Thu Thuy, Nguyen Linh Trang, Dang Minh Huong Giang, Chu Thi Hue, Nguyen Thi Anh Huong, and Hoang Quoc Anh. A mini-review on analytical methods for polycyclic aromatic hydrocarbons in vietnamese food and beverage samples. *Vietnam Journal of Food Control*, 6(3):231–245, 2023.

[47] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. Vi-liquid: unknown liquid identification with your smartphone vibration. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 174–187, 2021.

[48] S. Jain, A. Mishra, W. Li, and L. Liu. Liquid: A wireless liquid identifier. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 201–212, 2021.

[49] Surabhi Jain, Nilesh K Tiwari, and M Jaleel Akhtar. Csiwc rf sensor for microfluidic non-contact quality assessment of milk. *International Journal of RF and Microwave Computer-Aided Engineering*, 32(2):e22962, 2022.

[50] SN Jha, P Jaiswal, MK Grewal, M Gupta, and R Bhardwaj. Detection of adulterants and contaminants in liquid foods-a review. *Crit Rev Food Sci Nutr*, 10(56):1662–84, 2016.

[51] W. Jia, H. Peng, N. Ruan, Z. Tang, and W. Zhao. Wifind: Driver fatigue detection with fine-grained wi-fi signal features. *IEEE Transactions on Big Data*, 6(2):269–282, 2020.

[52] Weijia Jia, Hongjian Peng, Na Ruan, Zhiqing Tang, and Wei Zhao. Wifind: Driver fatigue detection with fine-grained wi-fi signal features. *IEEE Transactions on Big Data*, 6(2):269–282, 2020.

[53] Zhu Juncen, Jiannong Cao, Yanni Yang, Wei Ren, and Huizi Han. mmdrive: Fine-grained fatigue driving detection using mmwave radar. *ACM Trans. Internet Things*, 4(4), November 2023.

[54] Sang-Joong Jung, Heung-Sub Shin, and Wan-Young Chung. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intelligent Transport Systems*, 8(1):43–50, 2014.

[55] Sinan Kaplan, Mehmet Amaç Güvensan, Ali Gökhan Yavuz, and Yasin Karalurt. Driver behavior analysis for safe driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):3017–3032, 2015.

[56] N. Keshari and S. P. Singh. Kaiser window based blind beamformers for radar application. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 12(1):1–8, 2019.

[57] Hee-Joung Kim and Donghoon Lee. Image denoising with conditional generative adversarial networks (cgan) in low dose chest images. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 954:161914, 2020.

[58] Seong-Kwan Ko and Yong-Soo Cho. High-efficiency super-resolution fmcw radar algorithm based on adaptive reduced sampling. *Sensors*, 18(6):1876, 2018.

[59] Wanzeng Kong, Lingxiao Zhou, Yizhi Wang, Jianhai Zhang, Jianhui Liu, and Shenyong Gao. A system of driving fatigue detection based on machine vision and its application on smart device. *Journal of Sensors*, 2015:548602:1–548602:11, 2015.

[60] Chia-Chen Kuo, Mao-Jiun Wang, and Jun-Ming Lu. Developing sizing systems using 3d scanning head anthropometric data. *Measurement*, 152:107264, 2020.

[61] Marc Lalonde, David Byrns, Langis Gagnon, Normand Teasdale, and Denis Laurendeau. Real-time eye blink detection with gpu-based sift tracking. In *Fourth Canadian Conference on Computer and Robot Vision (CRV'07)*, pages 481–487. IEEE, 2007.

[62] D. Lee, S. Oh, S. Heo, and M. Hahn. Drowsy driving detection based on the driver's head movement using infrared sensors. In *2008 Second International Symposium on Universal Communication*, pages 231–236, 2008.

[63] Dongwook Lee, Seungwon Oh, Seongkook Heo, and Minsoo Hahn. Drowsy driving detection based on the driver's head movement using infrared sensors. In *2008 Second International Symposium on Universal Communication*, pages 231–236. IEEE, 2008.

[64] H. Lee, T. H. Kim, J. W. Choi, and S. Choi. Chirp signal-based aerial acoustic communication for smart devices. In *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2407–2415, 2015.

[65] Artem A Lenskiy and Jong-Soo Lee. Driver's eye blinking detection using novel color and texture segmentation algorithms. *International journal of control, automation and systems*, 10(2):317–327, 2012.

[66] Lingling Li, Yangzhou Chen, and Zhenlong Li. Yawning detection for monitoring driver fatigue based on two cameras. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6. IEEE, 2009.

[67] Weitao Li, Yuyong Xiong, Cunju Wei, Saisai Chen, Wei Fan, and Zhike Peng. A robust and widely applicable compensation method for quadrature imbalance of doppler radar. *IEEE Transactions on Microwave Theory and Techniques*, 72(6):6507–6517, 2024.

[68] Zilin Li, Haixing Liu, Chi Zhang, and Guangtao Fu. Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. *Environmental Science and Ecotechnology*, 14:100231, 2023.

[69] Xiaolin Liang, Yongling Jiang, and Thomas Aaron Gulliver. An improved sensing method using radio frequency detection. *Physical Communication*, 36:100763, 2019.

[70] Y. Liang et al. Fg-liquid: A contact-less fine-grained liquid identifier by pushing the limits of millimeter-wave sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–27, 2021.

[71] Yumeng Liang, Anfu Zhou, Huanhuan Zhang, Xinzhe Wen, and Huadong Ma. Fg-liquid: A contact-less fine-grained liquid identifier by pushing the limits of millimeter-wave sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–27, 2021.

[72] X. Lin, H. Wu, G. Huang, Q. Wu, and Z.-P. Yao. Rapid authentication of red wine by maldi-ms combined with dart-ms. *Analytica Chimica Acta*, 1283:341966, 2023.

[73] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. Blinklistener: " listen" to your eye blink using your smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–27, 2021.

[74] Jian Liu, Lei Zhang, Zengping Xu, Jingguo Wei, and Hongyan Yan. Convex optimization based window design for radar imaging. *Sensors*, 20(15):4216, 2020.

[75] Sannyuya Liu, Liang Zhao, Xidong Yang, Yiming Du, Menglin Li, Xiaoliang Zhu, and Zhicheng Dai. Remote drowsiness detection based on the mmwave fmcw radar. *IEEE Sensors Journal*, 22(15):15222–15234, 2022.

[76] Lin Lu, Xujun Ma, Yue Liang, Zhiqiang Liu, Xiangning Fan, and Lianming Li. A 60-ghz hybrid fmcw-doppler radar for vibration detection with a robust i/q calibration method. *IEEE Sensors Journal*, 22(20):20464–20474, 2022.

[77] Q. Marashdeh, L.-S. Fan, B. Du, and W. Warsito. Electrical capacitance tomography-a perspective. *Industrial & Engineering Chemistry Research*, 47(10):3708–3719, 2008.

[78] Qussai Marashdeh, Warsito Warsito, Liang-Shih Fan, and Fernando L Teixeira. A multimodal tomography system based on ect sensors. *IEEE Sensors Journal*, 7(3):426–433, 2007.

[79] A. Mashko. Review of approaches to the problem of driver fatigue and drowsiness. In *2015 Smart Cities Symposium Prague (SCSP)*, pages 1–5, 2015.

[80] Alina Mashko. Review of approaches to the problem of driver fatigue and drowsiness. In *2015 Smart Cities Symposium Prague (SCSP)*, pages 1–5. IEEE, 2015.

[81] Jorg JM Massen, Kim Dusch, Omar Tonsi Eldakar, and Andrew C Gallup. A thermal window for yawning in humans: yawning as a brain cooling mechanism. *Physiology & Behavior*, 130:145–148, 2014.

[82] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko. Iot, big data, and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal*, 9(9):6305–6324, 2022.

[83] Bahram Mohammadi, Mahmood Fathy, and Mohammad Sabokrou. Image/video deep anomaly detection: A survey. *arXiv preprint arXiv:2103.01739*, 2021.

[84] P. M. Morse, R. H. Bolt, and R. L. Brown. Acoustic impedance and sound absorption. *The Journal of the Acoustical Society of America*, 12(2):217–227, 1940.

[85] S Nordbakke and Fridulv Sagberg. Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(1):1–10, 2007.

[86] M. Raja et al. 3d head motion detection using millimeter-wave doppler radar. *IEEE Access*, 8:32321–32331, 2020.

[87] Muneeba Raja, Zahra Vali, Sameera Palipana, David G Michelson, and Stephan Sigg. 3d head motion detection using millimeter-wave doppler radar. *IEEE Access*, 8:32321–32331, 2020.

[88] M Rasekh, Bhagyashree Puranik, Upamanyu Madhow, and M Rodwell. In-the-field calibration of all-digital mimo arrays. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1389–1394. IEEE, 2022.

[89] J. Saeedi and K. Faez. Synthetic aperture radar imaging using nonlinear frequency modulation signal. *IEEE Transactions on Aerospace and Electronic Systems*, 52(1):1–12, 2016.

[90] Shourjya Sanyal and Koushik Kumar Nundy. Algorithms for monitoring heart rate and respiratory rate from the video of a user's face. *IEEE Journal of translational engineering in health and medicine*, 6:1–11, 2018.

[91] Rubina Sarki, Khandakar Ahmed, Hua Wang, Yanchun Zhang, and Kate Wang. Automated detection of covid-19 through convolutional neural network using chest x-ray images. *Plos one*, 17(1):e0262052, 2022.

[92] Chris Schwarz, Hunza Zainab, Soura Dasgupta, and Justin Kahl. Heartbeat measurement with millimeter wave radar in the driving environment. In *2021 IEEE Radar Conference (RadarConf21)*, pages 1–6. IEEE, 2021.

[93] Khawla Seddiki, Frédéric Precioso, Melissa Sanabria, Michel Salzet, Isabelle Fournier, and Arnaud Droit. Early diagnosis: End-to-end cnn-lstm models for mass spectrometry data classification. *Analytical Chemistry*, 95(36):13431–13437, 2023.

[94] Khawla Seddiki, Philippe Saudemont, Nina Precioso, Maxence Wisztorski, Michel Salzet, Isabelle Fournier, and Arnaud Droit. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nature communications*, 11(1):5595, 2020.

[95] Argha Sen, Anirban Das, Prasenjit Karmakar, and Sandip Chakraborty. mmassist: Passive monitoring of driver's attentiveness using mmwave sensors. In *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 545–553. IEEE, 2023.

[96] Mojtaba Shamsipur, Najmeh Yazdanfar, and Mahnaz Ghambarian. Combination of solid-phase extraction with dispersive liquid-liquid microextraction followed by gc-ms for determination of pesticide residues from water, milk, honey and fruit juice. *Food chemistry*, 204:289–297, 2016.

[97] Yuhong Shu, Yong Wang, Xiaobo Yang, and Zengshan Tian. An improved denoising method for eye blink detection using automotive millimeter wave radar. *EURASIP Journal on Advances in Signal Processing*, 2022(1):1–18, 2022.

[98] Mohamad-Hoseyn Sigari, Mahmood Fathy, and Mohsen Soryani. A driver face monitoring system for fatigue and distraction detection. *International journal of vehicular technology*, 2013, 2013.

[99] Mohamad-Hoseyn Sigari, Mahmood Fathy, and Mohsen Soryani. A driver face monitoring system for fatigue and distraction detection. *International journal of vehicular technology*, 2013, 2013.

[100] Gulbadan Sikander and Shahzad Anwar. Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6):2339–2352, 2019.

[101] Lee Slater and Andrew Binley. Advancing hydrological process understanding from long-term resistivity monitoring systems. *Wiley Interdisciplinary Reviews: Water*, 8(3):e1513, 2021.

[102] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[103] David Sommer and Martin Golz. Evaluation of perclos based current fatigue monitoring technologies. In *2010 annual international conference of the ieee engineering in medicine and biology*, pages 4456–4459. IEEE, 2010.

[104] Jim Stein. Sensing the road ahead. *Imaging and Machine Vision Europe*, 2017.

[105] Xue Sun, Wenwen Deng, Xudong Wei, Dingyi Fang, Baochun Li, and Xiaojiang Chen. Akte-liquid: Acoustic-based liquid identification with smartphones. *ACM Transactions on Sensor Networks*, 19(1):1–24, 2023.

[106] Xue Sun, Wenwen Deng, Xudong Wei, Dingyi Fang, Baochun Li, and Xiaojiang Chen. Akte-liquid: Acoustic-based liquid identification with smartphones. *ACM Transactions on Sensor Networks*, 19(1):1–24, 2023.

[107] Ye Sun, Xiong Yu, Jim Berilla, Zhen Liu, and Guangxi Wu. An in-vehicle physiological signal monitoring system for driver fatigue detection. In *3rd international conference on road safety and simulation purdue university transportation research board*, 2011.

[108] James L Szalma. Individual differences in stress, fatigue and performance. In *The handbook of operator fatigue*, pages 75–90. CRC Press, 2017.

[109] F. Tamm, G. Sauer, M. Scampicchio, and S. Drusch. Pendant drop tensiometry for the evaluation of the foaming properties of milk-derived proteins. *Food Hydrocolloids*, 27(2):371–377, 2012.

[110] Brian Christopher Tefft. Prevalence of motor vehicle crashes involving drowsy drivers, united states, 1999–2008. *Accident Analysis & Prevention*, 45:180–186, 2012.

[111] Gyungmin Toh, Seongwook Jeon, Junhong Park, and Won June Lee. Intraocular pressure measurement method using acoustic signals. *The Journal of the Acoustical Society of America*, 154(4_supplement):A137–A137, 2023.

[112] Udo Trutschel, Bill Sirois, David Sommer, Martin Golz, and Dave Edwards. Perclos: An alertness measure of the past. In *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 172–179, 2011.

[113] Alexander da Silva Vale, Cecília Marques Tenório Pereira, Juliano De Dea Lindner, Luiz Roberto Saldanha Rodrigues, Nájua Kêmil El Kadri, Maria Giovana Binder Pagnoncelli, Satinder Kaur Brar, Carlos Ricardo Soccol, and Gilberto Vinícius de Melo Pereira. Exploring microbial influence on flavor development during coffee processing in humid subtropical climate through metagenetic-metabolomics analysis. *Foods*, 13(12):1871, 2024.

[114] Anita N Vasavada, Jonathan Danaraj, and Gunter P Siegmund. Head and neck anthropometry, vertebral geometry and neck strength in height-matched men and women. *Journal of biomechanics*, 41(1):114–121, 2008.

[115] S Viswanathan, H Radecka, and J Radecki. Electrochemical biosensors for food analysis. *Monatsh Chem*, page 891–899, 2009.

[116] H. Wang, M. Xu, S. Li, and W. Cao. Mmliquid: Liquid identification using mmwave. In *2022 IEEE International Conference on Computer Communications (INFOCOM)*, pages 998–1008, 2022.

[117] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 288–300, 2017.

[118] Luping Wang, Yong Wei, Yuxiang Wang, Qiang Chen, Ping Liu, and Xiaofei Chai. Research on comprehensive and effective acoustic signal processing methods for caculating downhole liquid level depth. *Measurement*, 199:111452, 2022.

[119] Zhaoran Wang, Gilbert Lim, Wei Yan Ng, Pearse A Keane, J Peter Campbell, Gavin Siew Wei Tan, Leopold Schmetterer, Tien Yin Wong, Yong Liu, and Daniel Shu Wei Ting. Generative adversarial networks in ophthalmology: what are these and how can they be used? *Current opinion in ophthalmology*, 32(5):459–467, 2021.

[120] B. Warwick, N. Symons, X. Chen, and K. Xiong. Detecting driver drowsiness using wireless wearables. In *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, pages 585–588, 2015.

[121] Brandy Warwick, Nicholas Symons, Xiao Chen, and Kaiqi Xiong. Detecting driver drowsiness using wireless wearables. In *2015 IEEE 12th international conference on mobile ad hoc and sensor systems*, pages 585–588. IEEE, 2015.

[122] R. Wijasena and A. Munir. Optimization of weighting window functions for sar imaging via quadratically constrained quadratic programming. *Sensors*, 21(3):834, 2021.

[123] Jiramet Wongphanngam and Suree Pumrin. Fatigue warning system for driver nodding off using depth image from kinect. In *2016 13th International Con-*

ference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pages 1–6. IEEE, 2016.

[124] World Health Organization. *World Health Statistics 2017: Monitoring Health for the SDGs*. World Health Organization, 2017.

[125] Jincheng Wu, Ruixu Geng, Yadong Li, Dongheng Zhang, Zhi Lu, Yang Hu, and Yan Chen. Diffradar: High-quality mmwave radar perception with diffusion probabilistic model. In *ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8291–8295. IEEE, 2024.

[126] Y. Wu, P. Tian, Y. Cao, L. Ge, and W. Yu. Edge computing-based mobile object tracking in internet of things. *High-Confidence Computing*, 2(1):100045, 2022.

[127] Yalong Wu, Pu Tian, Yuwei Cao, Linqiang Ge, and Wei Yu. Edge computing-based mobile object tracking in internet of things. *High-Confidence Computing*, 2(1):100045, 2022.

[128] Binbin Xie, Jie Xiong, Xiaojiang Chen, Eugene Chai, Liyao Li, Zhanyong Tang, and Dingyi Fang. Tagtag: material sensing with commodity rfid. In *Proceedings of the 17th conference on embedded networked sensor systems*, pages 338–350, 2019.

[129] C. Yang, X. Wang, and S. Mao. Rfid-based driving fatigue detection. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2019.

[130] Chao Yang, Xuyu Wang, and Shiwen Mao. Rfid-based driving fatigue detection. In *2019 IEEE Global Communications Conference, GLOBECOM 2019, Waikoloa, HI, USA, December 9-13, 2019*, pages 1–6. IEEE, 2019.

[131] Chao Yang, Xuyu Wang, and Shiwen Mao. Respiration monitoring with rfid in driving environments. *IEEE Journal on Selected Areas in Communications*, 39(2):500–512, 2020.

[132] Hao Yang, Li Liu, Weidong Min, Xiaosong Yang, and Xin Xiong. Driver yawning detection based on subtle facial action recognition. *IEEE Transactions on Multimedia*, 23:572–583, 2021.

[133] Yanni Yang, Yanwen Wang, Jiannong Cao, and Jinlin Chen. Hearliquid: Nonintrusive liquid fraud detection using commodity acoustic devices. *IEEE Internet of Things Journal*, 9(15):13582–13597, 2022.

[134] Zheng Yang, Zhihan Li, Lei Song, Qi Wu, and Shan Fu. Evaluation research of joystick in flight deck based on accuracy and muscle fatigue. In Don Harris, editor, *Engineering Psychology and Cognitive Ergonomics - 11th International Conference, EPCE 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*, volume 8532 of *Lecture Notes in Computer Science*, pages 428–436. Springer, 2014.

[135] Shichao Yue and Dina Katabi. Liquid testing with your smartphone. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 275–286, 2019.

[136] Guangnan Zhang, Kelvin KW Yau, Xun Zhang, and Yanyan Li. Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis & Prevention*, 87:34–42, 2016.

[137] J. Zhang et al. Ubi-fatigue: Towards ubiquitous fatigue detection via contactless sensing. *IEEE Internet of Things Journal*, 2022.

[138] Jian Zhang, Yuan Wu, Yanjiao Chen, Junkongshuai Wang, Jinxing Huang, and Qian Zhang. Ubi-fatigue: Towards ubiquitous fatigue detection via contactless sensing. *IEEE Internet of Things Journal*, 2022.

[139] Linlong Zheng, Xiao Wang, and Marco Lops. Robust direction of arrival estimation using bayesian sparse learning under impulsive noise. *IEEE Transactions on Signal Processing*, 68:1951–1966, 2020.

[140] T. Zheng, Z. Chen, C. Cai, J. Luo, and X. Zhang. V2ifi: in-vehicle vital sign monitoring via compact rf sensing. *arXiv preprint arXiv:2110.14848*, 2021.

[141] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. V2ifi: in-vehicle vital sign monitoring via compact RF sensing. *CoRR*, abs/2110.14848, 2021.

[142] M. Zhu et al. Vehicle driver drowsiness detection method using wearable eeg based on convolution neural network. *Neural Computing and Applications*, 33(20):13965–13980, 2021.

[143] Miankuan Zhu, Jiangfan Chen, Haobo Li, Fujian Liang, Lei Han, and Zutao Zhang. Vehicle driver drowsiness detection method using wearable eeg based on convolution neural network. *Neural computing and applications*, 33(20):13965–13980, 2021.

[144] Łukasz Saganowski, Radosław Ośródka, Piotr Samczyński, and Marek Malinowski. Improvement in algorithms for quality control of weather radar data. *Atmospheric Measurement Techniques*, 15:261–277, 2022.