Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Surface-integrated sensing for human-robot interaction: methods, electronics, and validation in healthcare robotics

This is a PhD Thesis

Original Citation:
Surface-integrated sensing for human-robot interaction: methods, electronics, and validation in healthcare robotics / Leogrande, E.. - ELETTRONICO. - (2026).

Availability:
This version is available at http://hdl.handle.net/11589/304000 since: 2026-06-24

Published version
DOI:

Publisher: Politecnico di Bari

Terms of use:

(Article begins on next page)

21 July 2026

Elisabetta Leogrande

SURFACE-INTEGRATED SENSING FOR HUMAN-ROBOT INTERACTION: Methods, electronics, and validation in healthcare robotics

Thesis submitted for the degree of Philosophiae Doctor

Interuniversity Ph.D. Program in Industry 4.0
Polytechnic University of Bari

Tutors
Prof. Engr. *Francesco Dell'Olio*
Prof. Engr. *Stefano Mazzoleni*
Prof. Engr. *Luca De Cicco*

2026

The doctoral scholarship was funded by INPS – Direzione Regionale Puglia, CUP D93C23000050005.

Dissertation submitted for the degree of *Philosophiae Doctor*
Interuniversity Ph.D. Program in Industry 4.0

Cycle:
38th (2022, November 1 – 2026, April 30)

Administrative Headquarters: Politecnico di Bari – Department of Electrical and Information Engineering

Hosting University:
Eindhoven University of Technology – Department of Mechanical Engineering

Title:
SURFACE-INTEGRATED SENSING FOR HUMAN-ROBOT INTERACTION: Methods, electronics, and validation in healthcare robotics

Ph.D Candidate:
Elisabetta Leogrande

Academic Discipline:
Electronics (Italian code: IINF-01/A)

Tutors:
Prof. Engr. Francesco Dell'Olio, Politecnico di Bari (Bari, Italy)
Prof. Engr. Stefano Mazzoleni, Politecnico di Bari (Bari, Italy)
Prof. Engr. Luca De Cicco, Politecnico di Bari (Bari, Italy)

Coordinator:
Prof. Engr. Salvatore Digiesi, Politecnico di Bari (Bari, Italy)

External Reviewers:
Prof. Nunzio Cennamo, Università della Campania "Luigi Vanvitelli" (Caserta, Italy)
Prof. Armando Ricciardi, Università degli Studi Del Sannio (Benevento, Italy)

Last version:
May 5th, 2026

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program
SSD: IINF-01/A-Electronics

Final Dissertation

SURFACE-INTEGRATED SENSING FOR HUMAN-ROBOT INTERACTION: Methods, electronics, and validation in healthcare robotics

by
Elisabetta Leogrande

Supervisors:

Prof. Francesco Dell'Olio

Prof. Stefano Mazzoleni

Prof. Luca De Cicco

Coordinator of Ph.D. Program:
Prof. Salvatore Digiesi

Course n°38, 01/11/2022-30/04/2026

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program
SSD: IINF-01/A-Electronics

Final Dissertation

SURFACE-INTEGRATED SENSING FOR HUMAN-ROBOT INTERACTION: Methods, electronics, and validation in healthcare robotics

by

Elisabetta Leogrande

Elisabetta Leogrande

Referees:

Prof. Nunzio Cennamo

Prof. Armando Ricciardi

Supervisors:

Prof. Francesco Dell’Olio Francesco Dell' Olio

Prof. Stefano Mazzoleni

Prof. Luca De Cicco

Coordinator of Ph.D Program:

Prof. Salvatore Digiesi

Salvatore Digiesi

Course n°38, 01/11/2022-30/04/2026

1 Introduction

1.1 Motivation and Context

Healthcare systems increasingly face a dual challenge: the demand for care is rising, while clinical resources and time remain limited. Robots, especially collaborative robots designed to share space with humans, are often proposed as a way to support healthcare professionals in tasks that are repetitive, physically demanding, or require consistent monitoring.

Yet a robot deployed in a human-centered setting is only as useful as its ability to perceive people and the surrounding environment, and to translate that perception into safe, understandable, and effective behavior. Electronic skins (e-skins) are increasingly framed as enabling technologies for interactive human-machine interfaces, including in medical robotics scenarios[1].

A recurring limitation in current deployments is that perception is frequently concentrated in a small number of sensing points (e.g., internal joint sensing, end-effector force sensing, external vision). This sensing arrangement can be adequate for structured industrial tasks, but it becomes fragile in healthcare environments where contact is often incidental, surfaces are cluttered, human motion is unpredictable, and the cost of misinterpretation is high. Tactile sensation is explicitly highlighted as crucial for cobots to operate safely and effectively in dynamic, human-centric settings[1].

The motivation of this thesis is that robot surfaces themselves, the regions most likely to encounter people, can become meaningful sensing interfaces, enabling new forms of interaction and safety while also supporting physiological monitoring when appropriate. This direction is consistent with the view of e-skins as a convergence of materials science, sensor technology, and AI, spanning robotics and healthcare-oriented devices[2]–[4].

This work therefore addresses surface-integrated sensing as a unifying theme: (i) surface-like sensing for non-invasive monitoring of physiological variables and (ii) electronic skin solutions for tactile and proximity perception on robots, with an emphasis on electronics, signal processing, learning-based interpretation, and validation in scenarios relevant to healthcare robotics[5]–[8].

1.1.1 Human-Robot Collaboration in Healthcare Environments

Human-robot collaboration in healthcare differs from many industrial collaboration settings in at least three aspects. First, the interaction partner is not a trained operator alone: patients, clinicians, and caregivers have different expectations, behaviors, and tolerance to robot motion. Second, the environment is dynamic and space-constrained: equipment, beds, and wearable devices create frequent occlusions and unplanned contacts. Third, safety is not only a matter of collision avoidance; it also includes comfort, trust, hygiene constraints, and the ability to detect subtle interaction cues (e.g., light touch, approach, or patient motion).

In these contexts, sensing modalities that provide distributed, local, and immediate information are particularly valuable. The e-skin literature motivates tactile sensing for collaborative robotics due to safety, precision, and robustness in shared environments[1]. A robot that can detect an approaching limb, localize contact on its body, and estimate interaction force can enable safer physical collaboration, more natural handovers, and more robust operation near vulnerable users. Concrete system-level examples combine large-area multimodal tactile skins with local computing to support real-time interaction on robot surfaces[7].

The same distributed sensing philosophy also aligns with the trend toward continuous physiological monitoring using body-interfaced sensing technologies, which aim to capture clinically relevant signals without disrupting daily activities. Recent reviews frame e-skin and epidermal platforms as enablers for wearable/on-body monitoring and vital-sign sensing[9]–[11].

1.1.2 Limitations of Internal Robot Sensing and Control

Modern collaborative robots already include accurate internal sensing (joint encoders, motor currents, torque estimates) and control strategies (impedance/admittance, collision detection thresholds). However, internal sensing alone cannot reliably resolve where on the robot an interaction occurs, nor can it easily discriminate between different physical events (e.g., a soft incidental brush versus a constrained collision near a joint). These gaps are amplified when a robot is covered by accessories, compliant materials, or custom end-effectors, common in healthcare prototypes, because the “true” interaction point may be far from internal sensing references.

Moreover, relying on a single sensing modality can force overly conservative behavior. For example, collision detection based primarily on torque residuals may require high thresholds to avoid false positives during nominal motion, and that can delay detection of light contact. Similarly, camera-based perception can fail under occlusions, lighting changes, or privacy constraints. The result is often an engineering compromise: either the robot becomes cautious to the point of reduced usefulness, or it remains efficient but with reduced interaction awareness.

These limitations motivate a different sensing allocation: shifting part of perception from inside the robot to its external surfaces, where interaction information is physically generated. In this direction, the literature stresses that tactile sensing needs to be coupled with algorithms and control strategies (including ML and fusion approaches) to convert surface signals into meaningful interaction responses[8], [12].

1.1.3 Toward Surface-Based Sensing for Human-Robot Interaction

Surface-based sensing aims to make the robot body an active interface rather than a passive structure. Electronic skins, tactile arrays, and proximity-sensitive surfaces are examples of this direction, enabling detection of contact location, force distribution, deformation, or approach without requiring line-of-sight. A key insight is that such systems are not only sensor layers: they are electromechanical systems whose performance depends on materials, transduction principles, electronics, and signal interpretation pipelines[2].

At the same time, surface-integrated sensing is also central to wearable and epidermal technologies for physiological monitoring. Here, the “surface” is the body: signals such as photoplethysmography can provide access to cardiovascular information in a comfortable and continuous manner. In the literature, this perspective is connected to e-skin technologies for wearable monitoring and vital-sign-oriented platforms[5], [6], [9]–[11].

Despite different application targets (robot safety vs health monitoring), both domains share common methodological challenges: low-power electronics, robustness to motion and variability, signal quality assessment, and learning models that generalize across users and conditions. Recent reviews emphasize ML and multimodal fusion for perception and decision-making in e-skin systems and healthcare-oriented sensing pipelines[8], [12].

This thesis adopts surface-integrated sensing as a bridge between these domains, with the goal of contributing methods and validated implementations that are relevant to healthcare robotics.

1.2 Research Objectives

The thesis is guided by the following objectives:

I. **Design and validate surface-compatible sensing solutions** that can operate reliably in human-centered environments, where variability, soft contact, and motion artifacts are expected rather than exceptional[1], [7].

II. **Develop embedded and edge-oriented electronics and processing pipelines** enabling real-time operation with constrained power, computation, and communication resources[7], [13].

III. **Formulate learning-based methods** that translate raw surface signals into actionable quantities (e.g., physiological parameters, touch localization, interaction features), while addressing generalization and data quality[8], [12].

IV. **Demonstrate and evaluate the proposed systems** through experimental protocols and validation procedures aligned with healthcare-relevant use cases and constraints[14].

1.3 Contribution of the Thesis

This thesis contributes to the broader goal of enabling robots and devices to interact with humans through sensing that is distributed, compliant, and integrated into surfaces[1], [2], [8]. Contributions are grouped into two application streams and a set of cross-cutting methodological results.

1.3.1 Contributions to Surface-Integrated Vital Sign Monitoring

The first contribution stream targets non-invasive physiological monitoring using surface-compatible sensing. The thesis develops a workflow spanning sensor acquisition, preprocessing, dataset construction, model design, and evaluation for estimating cardiovascular quantities from optical signals. A key emphasis is placed on practical deployment constraints: robustness to subject variability and motion, model compactness, and suitability for embedded inference[5], [6], [9]–[11], [13].

1.3.2 Contributions to Electronic Skin for Human-Robot Interaction

The second contribution stream addresses robotic e-skin for safe and effective physical interaction. The thesis proposes and validates e-skin sensing elements and electronics capable of capturing interaction-relevant signals (including contact/proximity cues where applicable) and introduces algorithms to infer meaningful interaction descriptors such as location and force-related quantities. The work also examines how sensing architecture choices (materials, transduction, sensor placement, and readout electronics) affect performance, robustness, and integration feasibility on robot bodies[7], [8].

1.3.3 Cross-Cutting and Methodological Contributions

Across both streams, the thesis contributes:

- Hardware-software co-design principles for surface-integrated sensing systems, connecting transducer behavior, readout electronics, and inference models[7], [8].
- Data-centric practices for building reliable datasets and evaluation protocols in settings where ground truth can be expensive or noisy[8], [12].
- Edge/embedded deployment considerations, including model compression, latency/throughput trade-offs, and on-device validation strategies[13].

1.4 Structure of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 introduces background concepts and related work on surface-integrated sensing, tactile/proximity technologies, and learning methods for interpreting surface signals, with a focus on challenges relevant to healthcare robotics. Chapter 3 presents the contributions on surface-integrated physiological monitoring, including sensing setup, datasets, model development, and validation. Chapter 4 presents the FBG-based electronic skin, its electronics and integration aspects, and the algorithm used for touch localization deployed on embedded electronics. Chapter 5 presents the capacitive-pneumatic electronic skin for proximity and 3D object localization, including hardware refinement, dataset construction, and simulation-assisted learning framework. Chapter 6 provides a comparative and system-level discussion across the three experimental systems, distilling shared methodological lessons and implications for healthcare robotics. Finally, Chapter 7 concludes the thesis with a summary of contributions, limitations, and future research directions.

Figure 1.1 summarizes this organization, showing how the three experimental systems instantiate the surface-integrated sensing paradigm and how they connect to the comparative and concluding chapters.

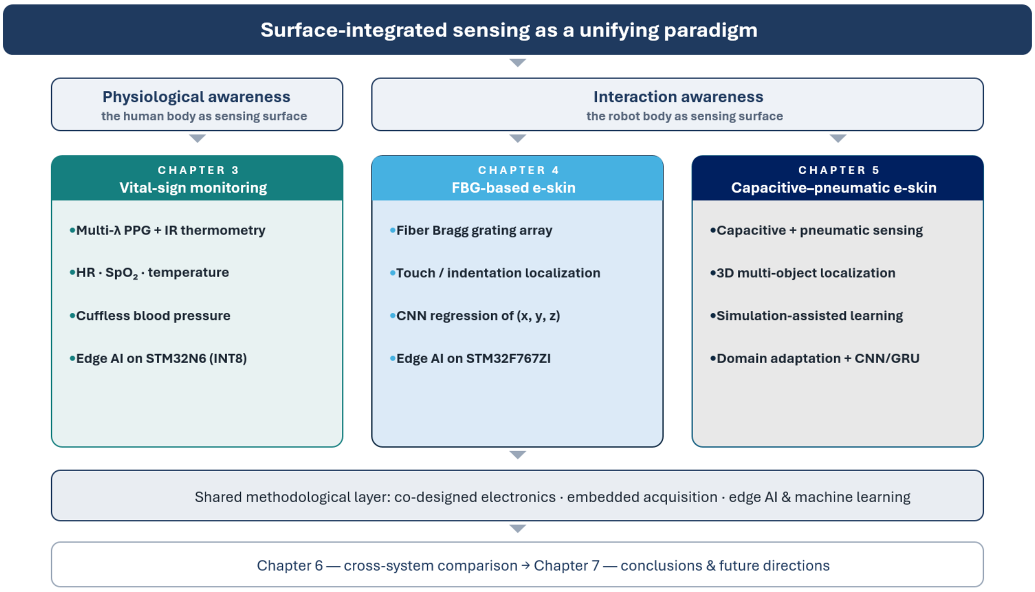

Figure 1.1. *Graphical overview of the thesis: surface-integrated sensing as a unifying paradigm bridging physiological and interaction awareness, instantiated by three experimental systems (Chapters 3-5) over a shared methodological layer and synthesized in Chapters 6-7.*

2 Surface-Integrated Electronics for Human-Robot Interaction

Chapter 1 framed surface-integrated sensing as a unifying direction for safer interaction and richer perception in healthcare robotics. Building on that motivation, this chapter consolidates the technical background needed to design sensing surfaces that are mechanically compatible with robot bodies, electrically reliable, scalable in wiring and computation, and interpretable through embedded processing. The focus here is on how surface-based systems are engineered, what constraints dominate their performance, and how acquisition/processing choices shape downstream learning and validation.

2.1 Background on Robot Sensing Solutions

Robots deployed near people typically rely on a layered sensing stack. Some modalities provide global information (e.g., scene geometry), others provide contact-level information (e.g., forces, deformations), and others only infer interaction indirectly (e.g., motor currents). A key design question for human-robot collaboration is not which sensor is "best," but how sensing is distributed across the robot and how quickly meaningful interaction cues can be extracted.

2.1.1 Sensing in Collaborative Robots

Collaborative robots usually combine proprioceptive sensing (joint encoders, torque/current estimates, internal observers) with exteroceptive sensing (vision, external force/torque sensors, safety scanners) and sometimes end-effector instrumentation (tool-mounted sensors). This stack is effective when interaction occurs at the tool tip or within anticipated regions. However, in many healthcare tasks, interaction is distributed along the robot body: incidental touch, constrained contacts near joints, or approach events that precede contact.

From a human-robot interaction (HRI) control perspective, the classical taxonomy of sensor-based interaction control emphasizes that sensing choice is inseparable from the control objective (force regulation, impedance shaping, constraint handling, etc.). In this view, a sensing surface can be interpreted as an additional measurement layer that reduces ambiguity about *where* interaction occurs and *how* it evolves over time, enabling more targeted controller responses[15].

A second viewpoint comes from robot skin surveys, where the core motivation is to extend perception from “a few sensing points” to distributed taxels that represent the robot body as an interface rather than a rigid shell. These works stress that the limiting factors are frequently system-level: wiring complexity, readout electronics, robustness, and the ability to run inference close to the sensing layer[16].

Finally, in collaborative robotics, sensing is coupled to safety concepts and deployment constraints. Surveys on collaborative robot safety highlight those real deployments balance responsiveness, nuisance triggers, and functional throughput; therefore, adding surface sensing is not automatically beneficial unless it can be integrated without compromising maintainability and reliability[17].

Taken together, these modalities cover a wide range of interaction cues, but their usefulness in close-contact collaboration depends on how quickly and unambiguously they reveal *where* and *how* an interaction is unfolding.

2.1.2 Safety, Adaptability, and Human Awareness

This is where safety and adaptability become system properties: they emerge from sensing, processing, and control timing acting as a single loop. In close-proximity HRI, “safety” is not a single metric. Besides limiting peak forces, systems must react early enough to avoid escalation, interpret ambiguous events (brush vs. sustained contact), and remain robust to environmental variability. Surface sensors contribute by increasing observability: they can directly measure local contact/proximity phenomena rather than inferring them indirectly from joint-level residuals.

A relevant line of work considers event-based and neuromorphic tactile sensing, motivated by the fact that high-density skins become communication- and compute-limited if they stream raw continuous data. Event-driven sensing aims to transmit only salient changes, which is attractive for large-area coverage[18].

Another perspective emphasizes embodied perception: tactile-rich bodies can support higher-level interpretation of contact patterns and environment properties, enabling interaction that is less brittle than line-of-sight sensing alone. An example is optoelectronic "robotic flesh" approaches that frame tactile perception as a data-rich sensing substrate for intelligent agents[19].

The practical question is therefore not whether surface sensing can add information, but whether it can be integrated in a way that remains reliable, serviceable, and scalable on real robot bodies.

2.2 Design Requirements for Surface-Based Systems

For that reason, the discussion now shifts from sensing concepts to design requirements, mechanical, electrical, and architectural, that ultimately determine whether surface perception can be deployed outside laboratory conditions[16].

Surface-based systems are constrained by the robot's geometry, materials, cabling and EMI environment, maintenance requirements, and the need to scale from prototypes to larger coverage. The constraints below are written to guide "design decisions": sensor, electronics, packaging, processing.

2.2.1 Mechanical and Conformal Constraints

A robotic surface is rarely a flat PCB-friendly platform: it includes curvature, seams, covers, and motion near joints. This creates three recurring requirements:

- **Conformability without performance collapse**. Flexible/soft layers must bend without changing baseline or sensitivity unpredictably. Reviews on touch-sensing highlight that material choice and micro-structuring can strongly affect response under repeated deformation and strain[20].

- **Durability under repeated contact.** Surface sensors face wear, compression set, and micro-damage. Self-healing and mechanically resilient materials are often proposed to extend lifetime and reduce maintenance. Representative materials-focused works include self-healing mechanisms and multifunctional hydrogel/organohydrogel approaches[21]–[24].
- **Integration with coverings and hygiene layers.** In healthcare contexts, surfaces may be covered (protective films, disposable covers) or cleaned frequently. Mechanically, this pushes designs toward layered stacks where sensing must remain stable despite added interfaces and thickness variations. Ultra-thin encapsulation materials are therefore not an "add-on," but part of the sensing chain[25].

In other words, the mechanical stack defines not only what can be mounted, but also the stability envelope within which signals can be interpreted.

2.2.2 Electronics Integration and Packaging

Once that envelope is fixed, the dominant sources of uncertainty often move to the electrical domain: parasitics, grounding, shielding, and packaging become first-order design variables. Surface sensing becomes an electronics problem as soon as the number of channels grows. Key points include:

- **Readout scalability and front-end design.** Large-area skins must multiplex channels, manage offsets and drift, and avoid crosstalk. Surveys on tactile arrays and driver ICs emphasize that array size often forces architectural choices (scanning strategies, shielding, local conditioning)[26].
- **Parasitics, EMI, and grounding on robots.** Robot bodies are electrically noisy environments (motors, switching drives, long cable runs). Packaging choices (flex PCB routing, ground planes, shielding layers) materially affect SNR and stability, especially for capacitive/proximity modalities.
- **Rigid-flex partitioning.** Many practical skins use "rigid islands" for electronics and "soft interconnects" for compliance. This reduces failure risk at solder joints and allows repairability by replacing a tile rather than the full skin.
- **Alternative transduction and optical readouts.** Optical tactile sensing provides immunity to electromagnetic interference and can be embedded into flexible

substrates. Pixel-based optical fiber sensing is a representative example of optical tactile force sensing for manipulation tasks[27].

Robust readout and packaging make surface signals measurable; scalability requires that they remain measurable as channel count grows.

2.2.3 Power, Communication, and Scalability

Scaling, however, is limited less by sensors than by power distribution, communication bandwidth, and where computation is placed in the system[28]. Surface systems face a three-way trade-off: power budget, data movement, and latency.

- **Power sourcing and harvesting.** When wired power is constrained or when designers aim to reduce tethering, self-powered approaches become attractive. Triboelectric nanogenerators are frequently discussed as flexible energy sources, and multiple works explore their use both for harvesting and as sensing mechanisms[29].
- **Communication architecture.** Centralized streaming of raw taxel data does not scale. Practical systems often adopt hierarchical buses: local tiles aggregate, preprocess, and transmit either compressed features or sparse events. This resonates with reference architectures discussed in broader industrial system design contexts (useful as a conceptual template for modularity and interoperability)[28].
- **Local compute placement.** If inference must be real-time, compute placement matters. Local processing reduces bandwidth and latency but requires careful model choice and validation on target hardware.

These constraints push designs toward modularity and local aggregation, which reduces data movement at the price of tighter node-level resource budgets.

2.2.4 Trade-Offs and Co-Design Strategies

The result is a co-design problem: mechanical choices affect drift and hysteresis, electronics affect noise and bandwidth, and processing choices affect latency and interpretability.

Surface-integrated sensing is inherently multi-objective. A useful way to structure decisions is to make explicit the coupled variables:

- **Transducer choice, readout electronics and signal representation.** For example, a design that chooses high-density taxels must either accept high data rates or adopt sparse/event representations.

- **Mechanical stack and calibration stability.** Softer layers can improve comfort and contact conformity but may increase drift/hysteresis; this shifts burden to compensation algorithms and periodic recalibration.
- **Model complexity and embedded feasibility.** Rich spatiotemporal models can boost performance but only help if they run within the system's power/latency envelope.

Recent work argues for end-to-end views where sensing, readout, and learning are considered jointly for high-density e-skin systems[30].

Similarly, deep learning surveys in tactile understanding highlight how representation choices (raw vs. learned features, spatiotemporal encoding) shape what interaction properties can be inferred robustly[31]. Making these trade-offs explicit is useful, but it is not sufficient: the system must still produce stable data and actionable outputs under realistic interaction conditions.

2.3 Data Acquisition and Embedded Processing

The remainder of this chapter therefore focuses on acquisition and embedded processing pipelines, how surface signals are collected, represented, interpreted, and validated under constraints that mirror deployment[32], [33]. Once a surface is physically integrated and electrically readable, performance is largely determined by acquisition protocol, preprocessing, representation, and how models are validated and deployed. This section provides a practical lens: what tends to break first in real data, and how to structure pipelines that remain interpretable and portable to embedded targets.

Figure 2.1 summarizes the generic acquisition and embedded-processing chain that the three experimental systems of this thesis instantiate; the co-design loop and the cross-cutting constraints highlighted at the bottom recur throughout Chapters 3-5.

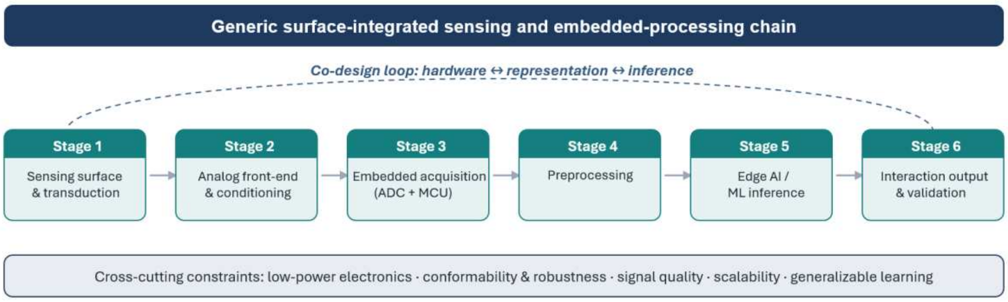

Figure 2.1. *Generic surface-integrated sensing and embedded-processing chain. Each experimental system in this thesis instantiates these stages; the co-design loop and the cross-cutting constraints recur across Chapters 3-5.*

2.3.1 Signal Characteristics and Noise Sources

Surface sensors exhibit characteristic distortions that differ across modalities:

- **Hysteresis and viscoelastic effects** (common in soft stacks) generate history dependence: the same force can produce different outputs depending on loading path.
- **Crosstalk** arises in dense arrays (electrical coupling, mechanical stress propagation), blurring localization.
- **Temperature sensitivity** affects many transducers and can mimic slow "drift."
- **Motion-induced artifacts** occur when the robot moves under the skin or when cabling flexes.

Tactile sensor comparisons consistently emphasize that each modality's "advantages" come with specific failure modes (e.g., capacitive: high resolution but parasitics/crosstalk; piezoresistive: low cost but temperature sensitivity; triboelectric: strong for dynamics but limited for static loads; optical: EMI immunity but size/cost)[26]. A useful engineering implication is that noise mitigation is not only filtering: it can require mechanical design choices (layering, encapsulation), electrical layout (shielding/grounding), and protocol design (repeatable contact trajectories, controlled loading rates). Because many disturbances are structured (mechanically and electrically), "noise handling" is as much about representation as it is about filtering.

2.3.2 Feature Extraction and Data Representation

Representation choices (raw streams, spatial maps, or sparse events) define what the model can learn and what the embedded platform can afford[18].

Surface signals can be represented at multiple abstraction levels:

1. **Taxel-level time series** (raw channels per location), suitable when arrays are small or when local inference is used.
2. **Spatial maps / tactile images** (taxel grid reshaped as an image), enabling convolutional models for localization and pattern recognition.
3. **Event streams** (changes-only), attractive for large areas or dynamic interaction where static streaming is wasteful.

For learning-based perception, the literature on tactile understanding repeatedly shows that representation is often the decisive factor: models trained on poorly structured signals may appear accurate in-distribution but fail under new materials, contacts, or trajectories[31].

In practice, representation choice should be aligned with the downstream control need:

- if the robot needs **fast protective reactions**, use low-latency features (thresholds, events, local gradients);
- if the robot needs **rich interaction descriptors** (contact type, multi-contact patterns), use spatiotemporal maps and models that capture dynamics.

In practice, a good representation compresses variability without discarding cues that matter for interaction.

2.3.3 Machine Learning and Embedded Intelligence

This raises the question of which models provide sufficient robustness while remaining measurable in terms of latency, memory footprint, and energy on the target device. Embedded intelligence for surface systems has three non-negotiables: **latency**, **power**, and **validation on target**.

- **Model compression and efficiency.** Compression is not only for memory footprint; it affects bandwidth (feature size) and sometimes robustness. A pragmatic overview of compression options (quantization/pruning/distillation) is often used as a toolkit for embedded deployment[34].
- **Benchmarking under realistic constraints.** TinyML benchmarking work highlights that "accuracy" alone is insufficient; models must be compared under hardware constraints (latency/energy) and measurement methodology must be consistent[35].
- **Validation on target device.** A recurrent failure mode is measuring performance

on a workstation and assuming it transfers. Guidance on evaluating ML models directly on target devices stresses that performance characterization must include the full deployment stack (preprocessing, I/O, runtime)[32].

A useful takeaway for this thesis is that embedded ML should be treated as part of the sensing system: if inference timing, buffering, and sampling jitter are not controlled, the "model" becomes a moving target.

Deployment constraints therefore become part of the scientific claim: if the model cannot be validated in the deployed stack, its apparent performance is incomplete.

2.3.4 Validation Methodologies in Collaborative Scenarios

Validation must then be designed to reflect how the system will be used, measuring not only estimation accuracy, but timing behavior, robustness, and failure modes in collaborative scenarios[36].

Validation for surface-based HRI should connect three layers:

I. **Sensor-level validation**: noise floor, drift, repeatability, hysteresis, response time, crosstalk characterization, environmental sensitivity. This yields limits on what the system can detect reliably.

II. **Perception-level validation**: localization error distributions, force estimation error vs. contact conditions, detection latency, false positive/negative rates under representative motion. Work that demonstrates real-time touch localization and inference on robotic platforms motivates reporting *both* spatial error and timing behavior.

III. **System-level validation in interaction loops**: how the robot reacts when perception is imperfect. For healthcare-oriented collaboration, protocols should include:
    - incidental touches during nominal motion,
    - approach detection (pre-contact),
    - constrained contacts near joints,
    - interaction with soft tissues/garments,
    - occlusions and clutter typical of clinical spaces.

In addition, because large-area skins enable richer perception, it becomes meaningful to validate higher-level capabilities (e.g., object recognition or tactile scene understanding) when relevant to the robot's task. Examples of robot-hand tactile sensing integrated for object recognition illustrate that tactile sensing can contribute beyond “contact yes/no”[37].

More recent robotic skin concepts also explore tomographic approaches to localize touch on compliant skins[38].

In summary, surface-integrated HRI systems are defined by a coupled pipeline: mechanical integration bounds signal stability, electronics determine measurability and scalability, and embedded processing converts raw surface measurements into timely, actionable descriptors.

The next three chapters apply this pipeline logic to the two thesis streams. Chapter 3 focuses on surface-compatible physiological monitoring, where motion artifacts and inter-subject variability shape both representation and evaluation choices. Chapters 4 and 5 focus on robotic e-skin for interaction, where distributed sensing and on-surface inference must meet latency and robustness requirements under contact-rich conditions. The shared engineering narrative is the same: surface sensing becomes practically useful only when integration, representation, and validation are co-designed around deployment constraints.

3 Surface-Integrated Device for Vital Parameter Monitoring

This chapter integrates and extends material previously published in “Smart combination of ECG and PPG signals: an innovative approach towards an electronic device for vital signs monitoring”, C. Botrugno et al.[39], “Contact-Less Body Temperature Monitoring by Infrared Camera: Accuracy Preliminary Assessment” E. Leogrande et al.[40], “Smart Combination of PPG Signal and IR Camera: A Multi-Parameter Device for Vital Signs Monitoring”, C. Botrugno et al.[41], "A multiwavelength photoplethysmography dataset with blood pressure and heart rate reference measurements”, E. Leogrande et al.[42] and “From PPG to Blood Pressure at the Edge: Quantization-Aware Architecture Selection and On-MCU Validation”, E. Leogrande et al.[43]. In comparison with the original publications, this chapter provides a unified presentation of the methodology, and a more detailed discussion of the experimental framework.

Continuous monitoring of vital parameters is a key enabling factor for safer and more responsive human-robot interaction in healthcare environments, where early detection of abnormal physiological trends can support timely intervention and reduce clinical workload. Digital health frameworks increasingly emphasize unobtrusive sensing, remote supervision, and scalable measurement pipelines, driving the integration of compact, embedded sensing modules into assistive and collaborative platforms[44], [45].

This chapter presents a multi-parameter monitoring approach designed around two sensing modalities: (i) multi-wavelength photoplethysmography (PPG) for cardiovascular and oxygenation metrics, and (ii) a compact infrared (IR) thermal camera for contactless skin temperature assessment.

Multi-wavelength PPG is leveraged to improve robustness against subject- and placement-dependent variability, while the IR thermal module supports contactless monitoring for improved comfort[39], [46].

The chapter details the hardware architecture, integration constraints, experimental protocol and dataset creation, signal-processing and machine-learning methods, and performance results for heart rate (HR), oxygen saturation (SpO_2), body temperature, and blood pressure (BP). Special attention is dedicated to the reference-device strategy, since HR ground truth differs across experimental campaigns (sphygmomanometer-derived HR when BP is also acquired, pulse-oximeter-derived HR when the focus is HR/SpO_2 only), which impacts comparability and discussion.

3.1 Background on Non-Invasive Vital Sign Monitoring

Non-invasive physiological monitoring aims to extract clinically meaningful information while minimizing discomfort, setup complexity, and measurement artifacts.

In the context of digital health, this translates into systems that are portable, cost-effective, and capable of producing reliable outputs with limited calibration overhead[44], [45].

Optical biosignals, especially photoplethysmography, have become central due to their ease of integration and the breadth of parameters that can be derived from waveform morphology and multi-wavelength absorption. The use of multiple wavelengths is increasingly considered a practical method to improve robustness, particularly under motion and heterogeneous tissue conditions[46].

Photoplethysmography (PPG) is an optical sensing technique that measures variations in transmitted or backscattered light caused by cyclic blood-volume changes in the microvascular bed during the cardiac cycle. In practice, PPG signals are acquired using a light-emitting diode (LED) source and a photodetector placed on the skin surface, either in transmission mode (detector opposite to the LED, typically at the fingertip or earlobe) or in reflectance mode (LED and detector on the same side, commonly used at the wrist and other wearable-compatible sites) (Fig 3.1).

The measured waveform is commonly described as the superposition of two components:

- a **pulsatile (AC) component**, primarily reflecting heartbeat-induced arterial blood volume changes and carrying timing/morphological information used for HR and hemodynamic indices;
- a **slowly varying (DC) component**, driven by the quasi-static optical attenuation of tissue and blood, and modulated by factors such as respiration, sympathetic activity, vasomotor tone, temperature, and sensor contact conditions.

Measurement quality depends strongly on sensor placement, coupling pressure, motion, and local perfusion.

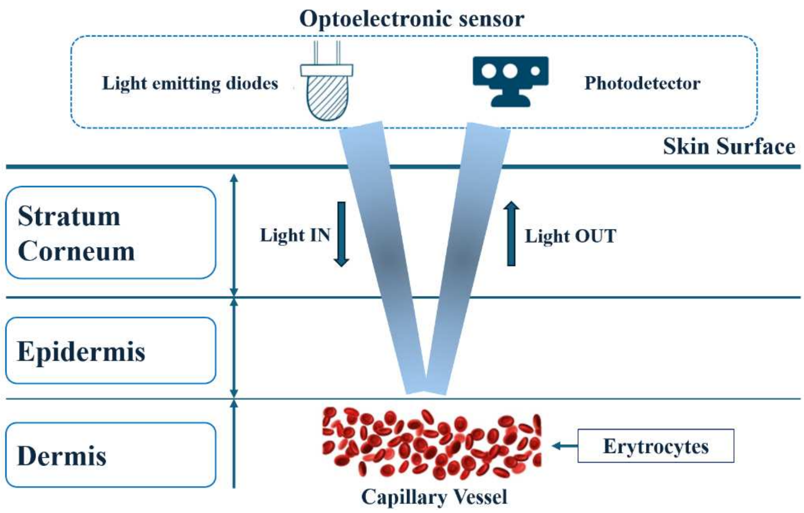

Figure 3.1. *Reflectance PPG working principle: LEDs illuminate the skin and the photodetector measures backscattered light whose intensity is modulated by pulsatile blood-volume changes in dermal microvasculature (AC component) superimposed on a slowly varying baseline (DC component).*

3.1.1 HR, SpO_2 and Temperature Monitoring

Heart rate (HR). HR can be derived from the pulsatile component of the PPG signal by identifying consecutive systolic peaks and computing the beat-to-beat period (RR interval). The quality of peak detection depends on sensor placement, contact pressure, motion, and wavelength-dependent interaction with tissue. In reflectance PPG, selecting a wavelength less prone to disturbances can improve peak detectability and reduce false detections[47].

Blood oxygen saturation (SpO_2). Pulse oximetry relies on wavelength-dependent absorption of oxygenated and deoxygenated hemoglobin, classically exploiting red and infrared channels. Multi-wavelength sensing can improve robustness under motion and provide additional degrees of freedom for artifact mitigation strategies[47], [48].

Temperature. Body temperature screening is clinically relevant for early identification of inflammatory and infectious states. From a device-design perspective, contactless or minimally-contact approaches are attractive for comfort, hygiene, and compatibility with fragile skin conditions. Systematic evidence comparing thermometer families supports the importance of controlled measurement conditions and device characterization[39].

3.1.2 Cuffless Blood Pressure Estimation

Cuffless BP estimation is typically approached through:

1. **physiology-driven models**, most notably based on pulse transit time (PTT) extracted from ECG and peripheral PPG;
2. **data-driven approaches**, where machine learning models regress SBP/DBP from features extracted from PPG only.

PTT-based methods exploit the inverse relationship between pulse wave velocity (and hence arterial stiffness) and transit time, often requiring robust detection of ECG R-peaks and PPG systolic peaks[49], [50].

Data-driven methods (e.g., neural networks) estimate BP directly from PPG temporal and spectral features, often aiming to reduce calibration needs and simplify hardware[51].

In addition to physiological and data-driven cuffless BP approaches, recent work has increasingly emphasized edge AI as a system-level requirement for wearable and surface-integrated monitoring. Executing inference directly on microcontroller-class hardware can reduce latency, limit continuous wireless streaming, and strengthen privacy by avoiding cloud dependence, but it introduces strict constraints on memory footprint, integer-only arithmetic, and real-time feasibility. In this perspective, it is no longer sufficient to report floating-point accuracy alone: the deployability of a BP estimator depends on how well the model preserves its behavior after full-integer quantization and how it performs when executed on the target microcontroller. These aspects are aligned with broader discussions in edge computing for healthcare and TinyML benchmarking, where hardware-dependent latency and memory cannot be reliably extrapolated from desktop evaluation[52], [53].

3.2 Hardware Architecture and Surface Integration

In surface-integrated vital monitoring, performance is ultimately constrained by the measurement interface: how sensing elements contact (or observe) the body, how stable geometry remains, and how acquisition/processing are implemented under embedded constraints. The hardware architecture is therefore described as a full chain (sensing front-end, electronics, mechanical interface, and integration rationale) so that protocols and results can be interpreted as consequences of concrete design choices rather than abstract sensing principles.

3.2.1 System Overview and Design Objectives

The design targets a multi-parameter device that can be integrated on an external surface (e.g., robot shell or dedicated module) and that supports:

- **Multi-parameter capability** using a minimal sensor set (PPG + thermal camera);
- **Non-invasive interaction**, enabling measurements with limited user burden;
- **Repeatable acquisition geometry**, especially for temperature (fixed distance, stable placement);
- **Low-cost prototyping and modularity**, enabling iterative integration and validation.

The architecture is organized into two subsystems: an optical PPG front-end for HR/SpO_2/BP and a thermal camera subsystem for contactless temperature.

3.2.2 Optical Sensing Front-End

The PPG subsystem is based on a reflectance multi-wavelength optical module, "MAX86916 EV System" (Fig. 3.2), integrating a photodiode and four LEDs with distinct spectral bands: red (655-663 nm), green (520-535 nm), blue (455-466 nm), and infrared (930-955 nm).

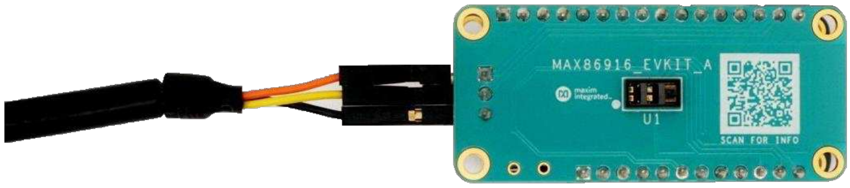

Figure 3.2. *MAX86916 EV System.*

The multi-wavelength output provides four synchronized PPG channels (Fig. 3.3) acquired during each session, enabling wavelength selection and artifact mitigation strategies.

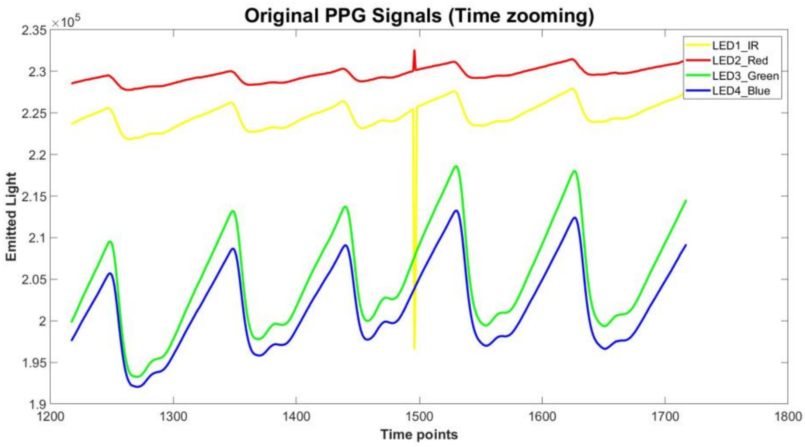

Figure 3.3. *Four PPG signals corresponding to four different wavelengths. IR and RED signals show rapid and wide peaks that should be removed.* (*x-axis: Time, y-axis: Emitted light in a.u.*).

3.2.3 Embedded Electronics and Processing Unit

The optical module is driven by an embedded microcontroller (Cortex-M4F class) that manages LED driving, acquisition timing, and data transmission. In the prototyping workflow, the evaluation system is connected to a host PC through USB interfaces (micro-USB plus a USB-FTDI link in the development setup), and data are streamed and logged through the vendor GUI into CSV files for offline processing.

For contactless temperature, the electronic system relies on a Teensy 4.0 microcontroller board (ARM Cortex-M7, up to 600 MHz, 2 MB flash and 1 MB RAM) interfaced via I²C to an MLX90640 thermal camera. The camera provides a 32×24-pixel array (768 pixels), and the embedded pipeline computes pixel temperatures from the raw frame data using the manufacturer’s conversion flow with emissivity set to 0.99. The acquisition chain includes serial communication toward the host PC (micro-USB link) for logging and evaluation (Fig. 3.4).

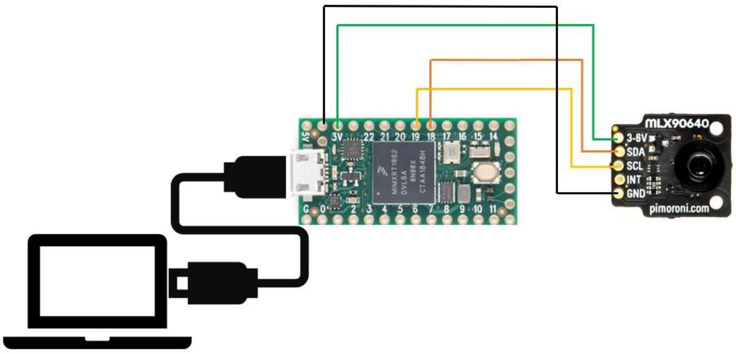

Figure 3.4. *Setup for body temperature.*

From an integration perspective, the temperature chain includes a dedicated mechanical constraint: a hollow, parallelepiped-shaped wrist support manufactured by 3D printing (Fig. 3.5). This fixture stabilizes distance and positioning, improving repeatability and reducing variance induced by hand posture and motion.

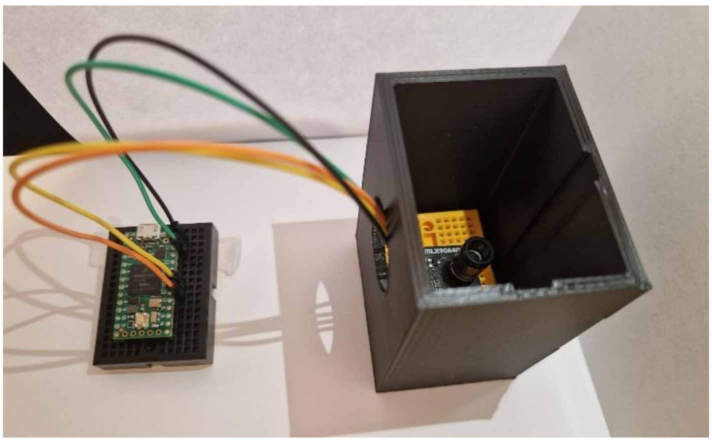

Figure 3.5. *Setup with the wrist support.*

To evaluate cuffless BP estimation under realistic embedded constraints, an additional processing path targets on-device inference on an STM32N6 microcontroller. In this workflow (Fig. 3.6), model development and dataset operations are performed offline: preprocessing and dataset preparation are implemented in MATLAB, while network training and validation are performed in TensorFlow (Google Colab). The selected model is exported as a TensorFlow Lite full-integer (INT8) network using post-training quantization with a representative calibration subset. The quantized TFLite model is then converted into optimized C code through the STM32Cube.AI toolchain, enabling integer-only execution on the STM32N6 target. The embedded deployment is evaluated using the toolchain analysis report (Flash/RAM occupation) and direct timing measurements on the microcontroller, ensuring that memory footprint and inference latency are explicitly accounted for as first-class design constraints in addition to estimation accuracy.

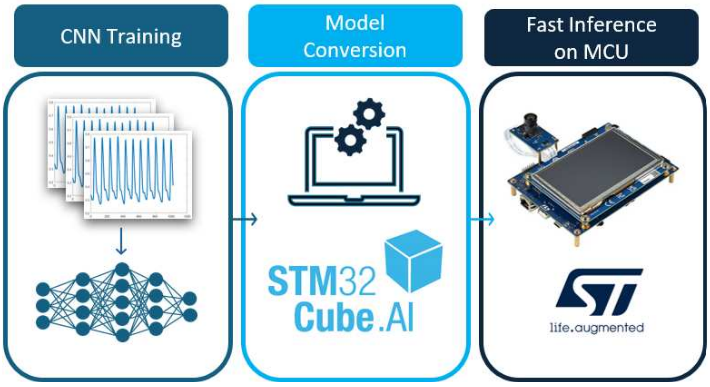

Figure 3.6. *Embedded deployment workflow.*

3.3 Experimental Protocol and Dataset Creation

This section describes the acquisition protocols used across the validation campaigns. A core point is to explicitly associate each campaign with its reference devices and to avoid mixing heterogeneous ground truths.

HR and SpO_2 campaign. For SpO_2 validation (and HR in the HR/SpO_2-focused acquisitions), a pulse oximeter provides the reference outputs. Multi-wavelength PPG is acquired simultaneously. The synchronous acquisition enables sample-aligned evaluation of estimated vs. reference values, while also exposing the pipeline to realistic motion artifacts that must be mitigated at preprocessing level.

BP campaign. For BP data creation, PPG is acquired from the right index finger while systolic and diastolic BP are measured on the left arm using a certified sphygmomanometer (OMRON, declared accuracy ±3 mmHg in the campaign) (Fig. 3.7). Each session lasts approximately 30s, consistent with the time required for oscillometric BP estimation stabilization. The campaign includes 88 individuals (balanced sex distribution) aged 19-84 years.

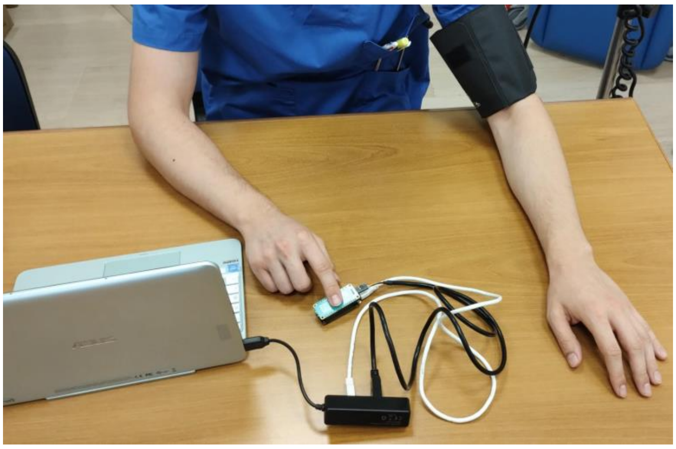

Figure 3.7. *Experimental set-up for blood pressure data acquisitions: PPG signals were acquired by positioning the right index finger on the PPG sensors and SBP and DBP via sphygmomanometer.*

Importantly, in this BP-oriented campaign, the HR reference is the sphygmomanometer readout, not the pulse oximeter. This should be treated as a distinct HR validation condition rather than pooled with pulse-ox-based HR results.

Temperature campaign. Temperature validation is performed using the MLX90640 thermal camera and a certified IR thermoscan as reference (declared accuracy ±0.2 °C). The wrist support is used to reduce distance and alignment variability. A cohort of 30 subjects aged 23-65 years is included in the preliminary assessment.

3.4 Signal Processing and Machine Learning Methods

HR estimation from PPG. HR is extracted using a single wavelength channel (infrared, around 950 nm in the HR-oriented pipeline). The raw IR signal is processed with a low-pass filter to remove high-frequency noise (Fig. 3.8).

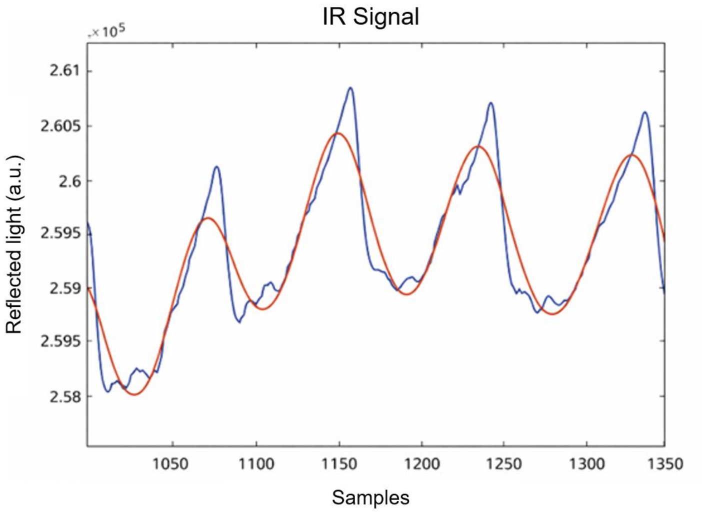

Figure 3.8. *IR signal before (blue signal) and after (red signal) filtering (x-axis: samples, y-axis: reflected light in a.u.).*

Peaks corresponding to consecutive beats are detected, and the RR interval is computed by dividing peak-to-peak distances (in samples) by the sampling frequency. HR in bpm is obtained as:

$$HR = \frac{60}{mean(RR)} \tag{1}$$

Performance is evaluated using absolute error between estimated and reference HR.

SpO_2 estimation from Red/IR PPG. SpO_2 estimation uses red and infrared PPG channels. The preprocessing pipeline includes: (i) moving-average filtering to remove outliers, followed by (ii) a cascade of low-pass and high-pass filters (band-pass effect). Since motion artifacts can strongly distort the signals, a PCA-based artifact reduction is applied by retaining components explaining 65% of the variance[54].

After artifact reduction, the root-mean-square (RMS) is computed for both red and infrared components:

$$RED = \sqrt[2]{\frac{\sum_{n=1}^{N} smooth\_red(n)^2}{N}} \tag{2}$$

$$IR = \sqrt[2]{\frac{\sum_{n=1}^{N} smooth\_ir(n)^2}{N}} \tag{3}$$

Then it is possible to compute the ratio R, as follows:

$$R = \frac{RED}{IR} \tag{4}$$

Once obtained the ratio R, the value of SpO_2 is obtained through a second-order calibration curve:

$$SpO_2 = K_1R^2 + K_2R + K_3 \tag{5}$$

where K_1, K_2 and K_3 are calibration coefficients, experimentally acquired from literature. SpO_2 performance is evaluated via relative percentage error.

Temperature estimation from thermal camera frames. For each thermal frame (32×24 pixels), object temperatures are computed from camera data with emissivity set to 0.99. A physiologically plausible temperature band is applied (35-45 °C): only pixels in-range are retained and averaged. This is repeated for 15 iterations (15 frames), and the final estimate is obtained as the mean across iterations. The workflow is shown below (Fig. 3.9).

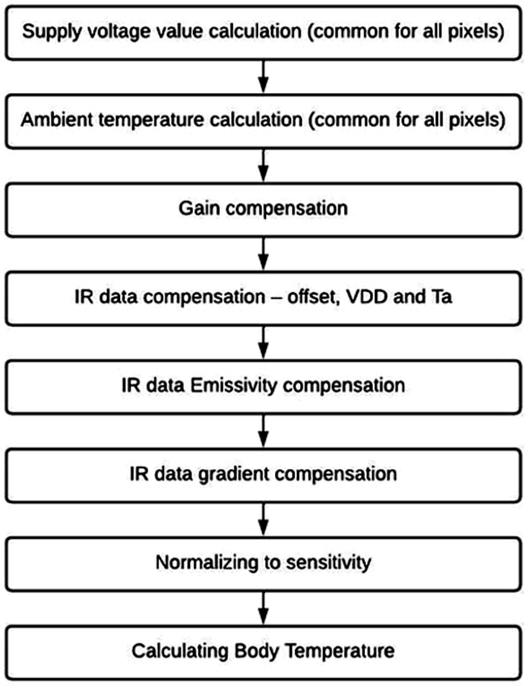

Figure 3.9. *Calculation Flow*

Blood pressure estimation. Three strategies are implemented:

I. **PPG-only BP via feed-forward ANN.** A feed-forward ANN estimates SBP and DBP from temporal and spectral features extracted from PPG after motion/slow-trend mitigation. The model is trained on a large PPG-BP database and tested on the newly acquired cohort, supporting regression of both SBP and DBP without requiring multiple sensors. Foundational neural-network approaches for continuous BP estimation from PPG have been reported in the literature[53].

II. **PTT-based BP via ECG+PPG.** ECG and PPG are filtered and processed through peak-detection. A Pan-Tompkins-style algorithm is adopted to detect ECG R-peaks (and corresponding complexes) and to support consistent timing extraction[50].

PTT is defined as the time difference between the ECG R-peak and the subsequent maximum PPG peak (Fig. 3.10).

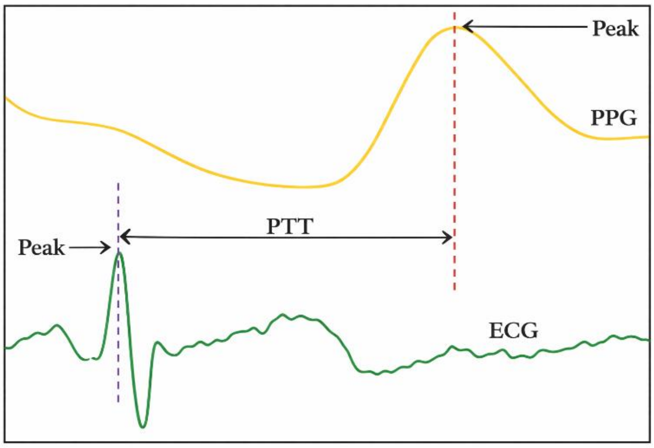

Figure 3.10. *Pulse Transit Time (PTT) as temporal distance between ECG R peak and PPG systolic peak (x-axis: time).*

The relationship between PTT and BP is consistent with prior approaches based on pulse transit time[55].

BP is estimated using the implemented model:

$$BP = \frac{A}{PTT^2} + B \tag{6}$$

where A is computed from the subject's height (with an assumed proportionality factor for the heart-to-fingertip path length) and B is obtained through a calibration procedure. In the implemented calibration, B is computed across different activity conditions (run, walk, sit) as $B_{run}, B_{walk}, B_{sit}$, and a weighted average prioritizes B_{sit}, since final estimation focuses on sitting segments.

III. **Edge-oriented PPG-only BP.** To explicitly address deployability on constrained hardware, an edge-oriented workflow was adopted to identify PPG-only neural architectures that retain BP estimation performance compatible with the AAMI/ESH/ISO thresholds considered in this study, after full-integer (INT8) quantization and during on-microcontroller execution. The study uses the public UCI "Cuff-Less Blood Pressure Estimation" dataset, where infrared PPG is paired with arterial blood pressure (ABP). After record selection and cleaning, PPG is segmented into 8.192 s windows (1024 samples at 125 Hz) with 75% overlap, extracting SBP/DBP as max/min of ABP in each window; PPG is detrended and min-max normalized to [0,1] (Fig. 3.11).

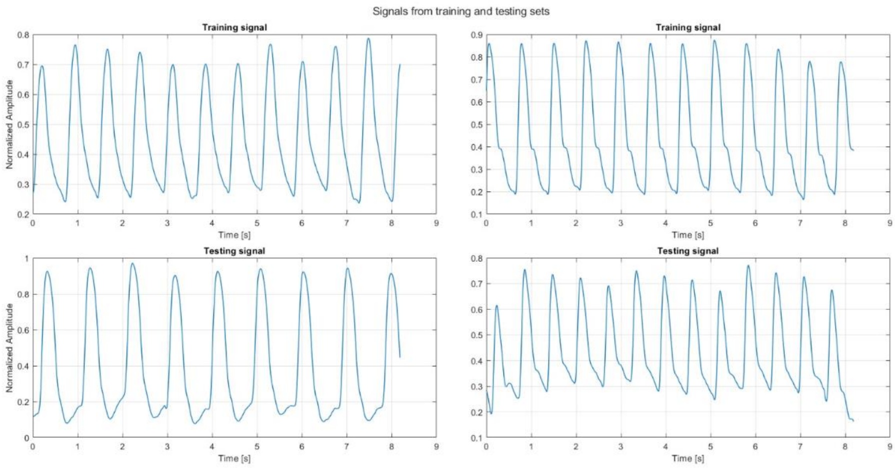

Figure 3.11. *Examples of normalized infrared PPG segments extracted from the UCI dataset, showing representative signals from both training (top) and testing (bottom) partitions*.

To avoid information leakage due to overlap, identical segments across splits are removed before final evaluation. Multiple 1D convolutional architectures are compared under a controlled protocol, spanning (i) a baseline 1D CNN, (ii) residual 1D CNN designs, including a compact residual “Slim” variant (reduced complexity with respect to the full residual model), (iii) MobileNet-1D based on depthwise separable convolutions, and (iv) a compact temporal convolutional network (Micro-TCN) (Table 3.1).

Table 3.1. *Architectural and training characteristics of the evaluated neural network models.*

Architecture	Input shape	Main architectural features	Loss	Optimizer	LR	Weight decay	Batch size	Training strategy
Baseline 1D CNN	(1024,1)	Stacked Conv1D + dense	MSE	Adam	1×10^{-3}	—	128	Fixed epochs (400)

		regression head						
Residual 1D CNN	(1024,1)	Residual blocks with increasing channels	Huber	AdamW	8×10^{-4}	1×10^{-4}	128	Early stopping
Residual 1D CNN (Slim)	(1024,1)	Compact residual architecture with GAP	Huber	AdamW	8×10^{-4}	1×10^{-4}	128	Early stopping
MobileNet-1D	(1024,1)	Depthwise separable 1D convolutions	Huber	AdamW	8×10^{-4}	1×10^{-4}	128	Early stopping
Micro Temporal Convolutional Network	(1024,1)	Dilated temporal convolutions in residual structure	Huber	AdamW	8×10^{-4}	1×10^{-4}	128	Early stopping

Training choices are harmonized to enable fair comparison (loss/optimizer, early stopping, batch size), and models are evaluated with metrics aligned with common validation practice for cuffless BP, including signed error mean (bias) and dispersion. Performance is reported and discussed with reference to the AAMI/ESH/ISO thresholds ($|ME| \le 5$ mmHg and $SD < 8$ mmHg, for both SBP and DBP). Crucially, the workflow includes a dedicated quantization stage: selected candidate models undergo full integer post-training quantization to INT8 (including INT8 input/output tensors), using a representative calibration subset to estimate activation ranges[56]. Quantization robustness is assessed by comparing error statistics before/after INT8 conversion and by analyzing prediction consistency (e.g., FP32 vs INT8 correlation). The most quantization-stable architecture is then exported in TensorFlow Lite INT8 format and passed through an industrial MCU toolchain to generate optimized C code for deployment on a microcontroller-class platform.

3.5 Experimental Results and Discussion

This section reports experimental outcomes and discusses their practical implications for healthcare-oriented human-robot interaction. Results are organized by physiological target to reflect the different sensing modalities and validation protocols: first, HR, SpO_2 and temperature are presented as primary multi-parameter outputs obtained from multi-wavelength PPG and the infrared thermal camera; then, blood pressure estimation is discussed through both data-driven PPG-only regression and physiology-driven timing-based estimation. In addition, an edge-oriented BP workflow is included to highlight how deployment constraints become part of the evaluation, beyond accuracy in floating-point offline testing.

The discussion therefore emphasizes not only the achieved accuracy, but also the conditions under which it is obtained, the main sources of variability, and the implications for real-world surface-integrated operation.

3.5.1 HR, SpO_2 and Temperature Estimation

HR Results. In the HR validation where the reference device is the sphygmomanometer (BP-oriented acquisition), the IR-channel low-pass filtering and peak-based RR extraction yield a maximum absolute error of 3 bpm, and multiple cases show zero difference between estimated and reference HR (Fig. 3.12). This indicates that, under controlled contact conditions, IR-channel selection combined with simple filtering can achieve reliable beat interval estimation.

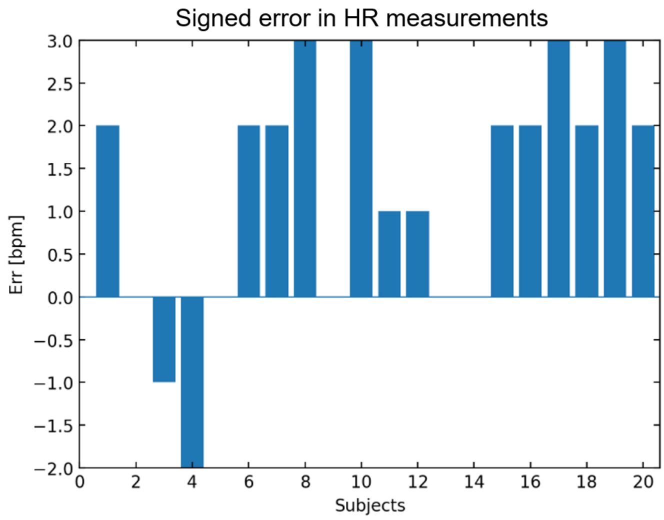

Figure 3.12. *Signed error in HR measurements across subjects (negative values indicate underestimation).*

SpO_2 Results. The combined preprocessing chain (moving average, band-pass filtering, and PCA retaining 65% variance) yields a maximum relative error around 5%, with multiple cases showing null error (Fig. 3.13). The polynomial mapping from ratio R to SpO_2 provides an embedded-friendly estimator; however, stability depends on maintaining a consistent optical stack and acquisition geometry aligned with the calibration source.

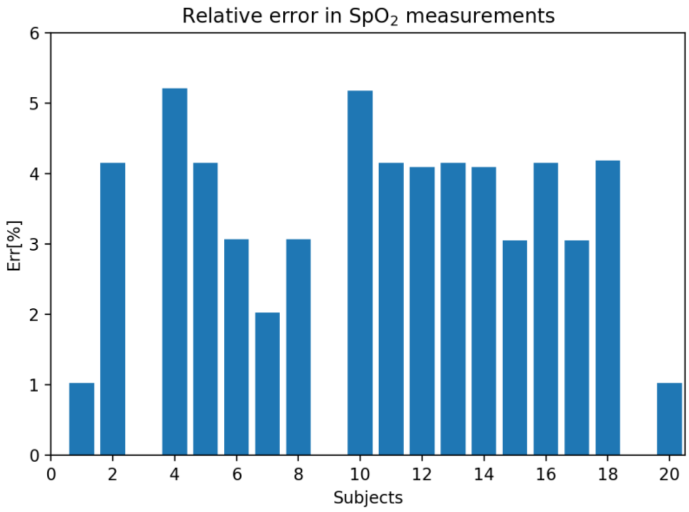

Figure 3.13. *Histogram shows relative errors for each subject.*

Temperature. The contactless temperature pipeline shows close agreement with the certified IR thermoscan reference (declared accuracy ±0.2 °C) under controlled acquisition conditions enabled by the 3D-printed wrist support, which stabilizes distance and alignment and reduces posture-induced variability. Object temperature is computed from MLX90640 raw frames using the manufacturer conversion flow (including supply/ambient compensation and gain/offset corrections) with emissivity set to 0.99. For each measurement, a physiologically plausible band (35-45 °C) is applied at pixel level and in-range pixels are spatially averaged; this estimate is repeated across 15 consecutive frames and the final value is obtained as the mean across iterations, reducing random pixel noise. Across the cohort, the mean absolute error is 0.11 °C (maximum absolute error 0.2 °C; minimum 0 °C), and the mean relative error is 0.29%. Residuals show an approximately Gaussian pattern in the Q-Q analysis (Fig. 3.14).

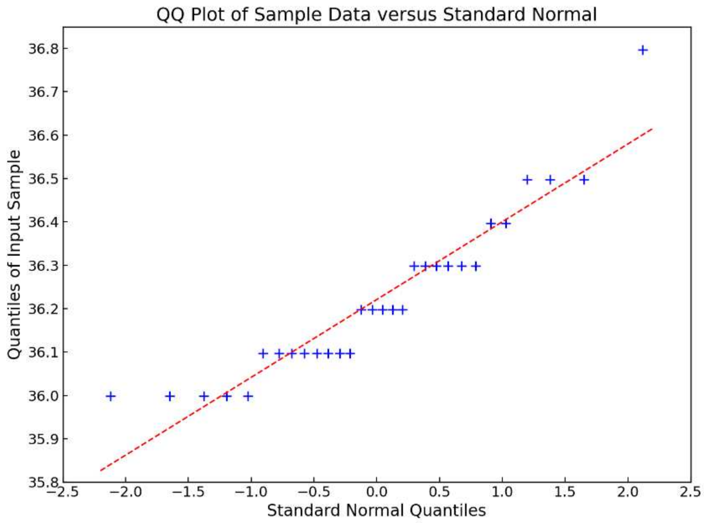

Figure 3.14. *Quantile-Quantile diagram. The red line describes the Gaussian distribution pattern, the blue crosses are the temperature samples acquired by IR camera MLX90640.*

Bland-Altman analysis shows a small negative bias (-0.08 °C) and 95% limits of agreement from -0.26 °C to +0.10 °C (MLX90640 - reference), indicating close agreement with few localized outliers (Fig. 3.15).

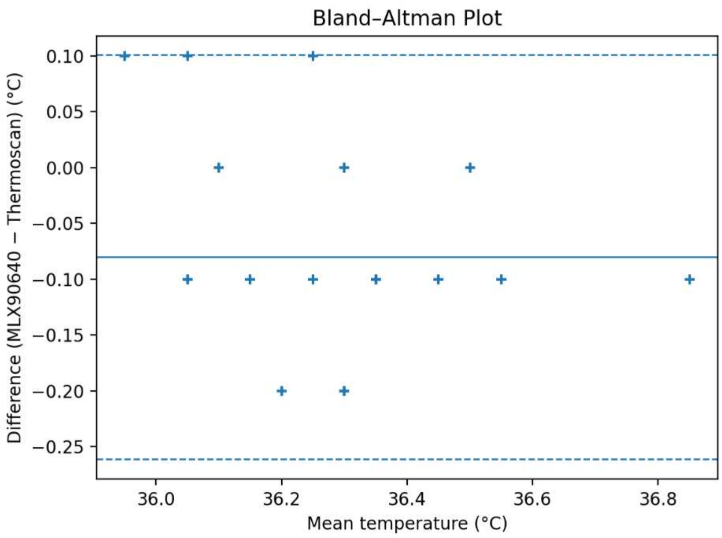

Figure 3.15. *Bland-Altman plot comparing MLX90640-based temperature estimates against the certified IR thermoscan reference. Solid line: mean difference (bias). Dashed lines: 95% limits of agreement (bias ± 1.96·SD).*

3.5.2 Blood Pressure Estimation

PPG-only ANN. On the BP cohort, the ANN-based regression achieves MAE of 5.08 ± 8.83 mmHg for SBP and 4.37 ± 7.08 mmHg for DBP. These results are promising for a prototypal workflow, while also highlighting that dispersion remains the critical factor to address for BP under inter-subject variability and artifact conditions.

PTT-based BP. Using ECG+PPG, the PTT-based estimator yields a maximum relative error below 5%, and multiple cases exhibit zero difference between estimated and reference BP (Fig. 3.16). This indicates that, for the public dataset condition, reliable peak detection is sufficient to support stable PTT extraction and physically grounded BP estimation, provided that activity-dependent calibration is handled explicitly.

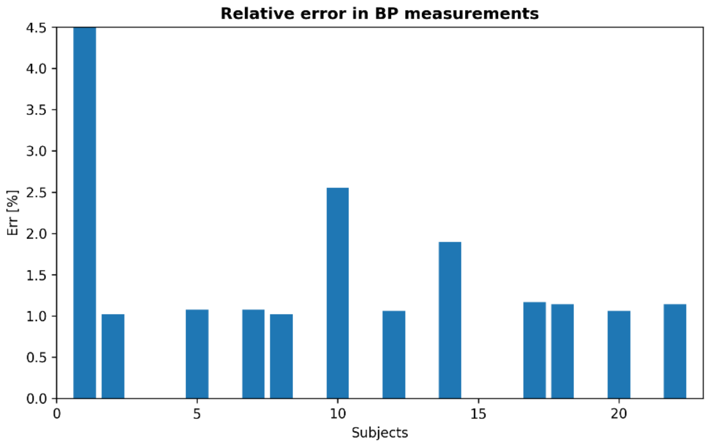

Figure 3.16. *Relative errors in the BP evaluation.*

Edge-oriented PPG-only BP. In the edge-oriented evaluation, five 1D CNN families are benchmarked on windowed infrared PPG segments extracted from the UCI dataset. The dataset preprocessing yields 129,276 windowed instances, split into 70% training and 30% test (with overlap-leakage removal). SBP spans approximately 78-200 mmHg, while DBP spans approximately 50-158 mmHg, resulting in a markedly skewed DBP distribution (Fig. 3.17).

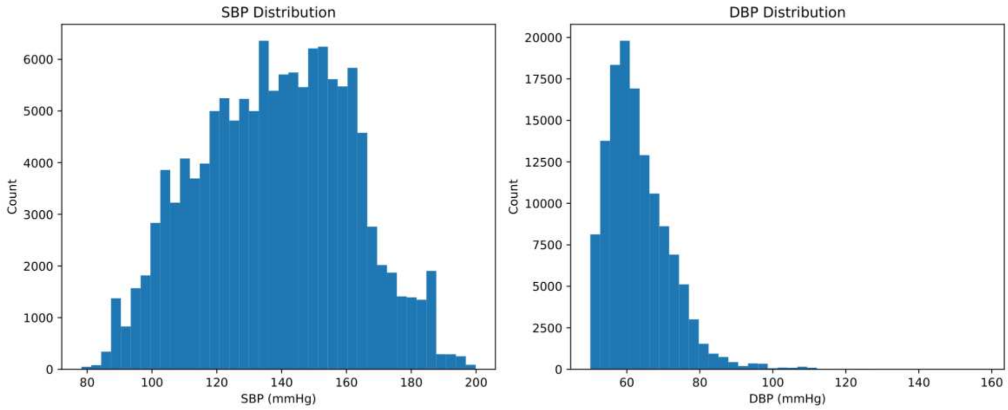

Figure 3.17. *Distribution of systolic and diastolic blood pressure values in the full dataset.*

Under floating-point inference, multiple architectures meet the AAMI/ESH/ISO error-dispersion thresholds on this dataset, but they show very different accuracy-complexity trade-offs (Table 3.2). Compact residual design achieves low dispersion while maintaining a moderate parameter count compared to larger residual networks.

A key finding is that floating-point accuracy is not a reliable predictor of deployability: the baseline CNN exhibits a noticeable degradation after full-integer INT8 conversion (bias amplification and increased variability, especially for SBP), whereas the compact residual “Slim” architecture preserves error statistics close to the floating-point case (Table 3.2) and maintains stable performance after quantization (Table 3.3). Importantly, the term memory reduction here refers to the reduction in weight storage when moving from FP32 to INT8 TensorFlow Lite representations: the Baseline model decreases from 137 KiB to 43 KiB (-68.6%), while the Residual 1D CNN-Slim decreases from 288 KiB to 166 KiB (-42.4%). Although the Slim model remains larger than the Baseline in absolute INT8 footprint (166 KiB vs 43 KiB), it provides a markedly better accuracy-robustness trade-off under integer-only inference, motivating its selection for on-device deployment.

Table 3.2. *Floating-Point blood pressure estimation performance.*

Architecture	SBP ME ± SD (mmHg)	SBP MAE (mmHg)	DBP ME ± SD (mmHg)	DBP MAE (mmHg)	# Params	Model size (KiB)
Baseline 1D CNN	-0.27 ± 7.38	5.13	0.01 ± 4.68	3.17	33,826	137
Residual 1D CNN	1.55 ± 3.35	2.59	0.36 ± 2.24	1.45	554,498	1094
Residual 1D CNN (Slim)	1.29 ± 4.57	3.30	0.75 ± 3.25	2.10	140,034	288
MobileNet-1D	0.12 ± 7.94	5.73	0.44 ± 4.39	3.04	10,738	37
Micro TCN	2.18 ± 7.05	5.42	0.20 ± 4.09	2.67	41,698	105

Table 3.3. *TFLite INT8 blood pressure estimation performance on the independent test set.*

Architecture	SBP ME ± SD (mmHg)	SBP MAE (mmHg)	DBP ME ± SD (mmHg)	DBP MAE (mmHg)	Model size (KiB)
Baseline 1D CNN	3.39 ± 8.20	6.47	3.24 ± 4.99	4.28	43
Residual 1D CNN (Slim)	0.62 ± 4.68	3.22	0.59 ± 3.26	2.07	166

The selected INT8 model is deployed on an STM32N6 microcontroller and validated under real execution constraints. The deployment report indicates Flash occupation on the order of hundreds of kilobytes and RAM usage below 100 KiB, with most RAM allocated to activations. On-device inference maintains low bias and limited dispersion, with mean errors within a few mmHg for both SBP and DBP (Fig. 3.18-3.19), while average inference time remains compatible with continuous operation given the physiological acquisition window.

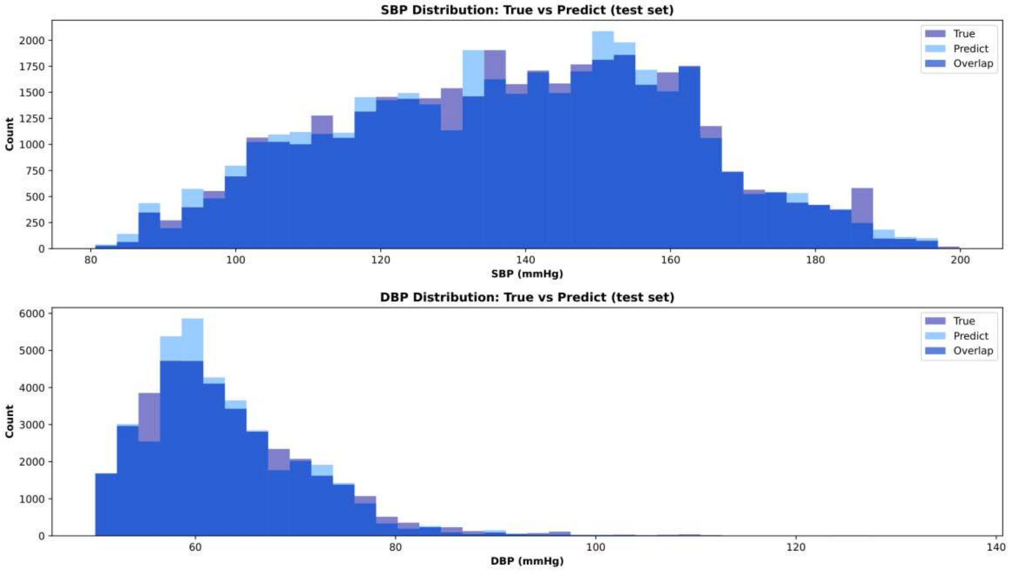

Figure 3.18. *Distribution of true and predicted systolic and diastolic blood pressure values obtained with the Residual 1D CNN-Slim model on the independent test set.*

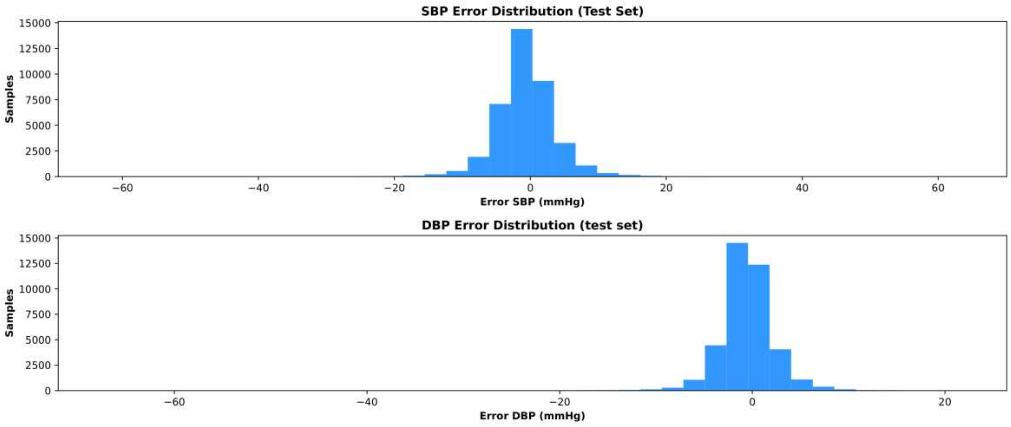

Figure 3.19. *Distribution of systolic and diastolic blood pressure estimation errors obtained with the Residual 1D CNN-Slim model on the independent test set.*

3.5.3 Limitations and Discussion

The validation presented in this chapter demonstrates the feasibility of multi-parameter monitoring with a surface-integrated sensing module. At the same time, several limitations must be stated explicitly to avoid over-generalization and to clarify which results can be compared directly.

First, methodological limitation concerns the heterogeneity of HR ground truth across experimental campaigns. When BP is acquired, HR reference values are taken from the sphygmomanometer readout, whereas in HR/SpO_2-focused sessions the HR (and SpO_2) reference is provided by a pulse oximeter. These instruments can differ in averaging windows, internal filtering, and artifact rejection; therefore, an aggregated HR metric across campaigns may introduce bias.

Regarding SpO_2, the estimation pipeline is compact and suitable for embedded implementation, but it remains sensitive to acquisition geometry and the optical stack. The polynomial mapping from the ratio R to SpO_2 implicitly assumes a stable relationship between Red/IR components and physiological oxygenation. Changes in sensor placement, contact pressure, ambient light leakage, or inter-subject tissue variability can shift this relationship. Multi-wavelength acquisition and PCA-based artifact reduction mitigate motion and noise, but they cannot fully eliminate the confounding effect of strong motion artifacts and poor contact conditions, especially in uncontrolled real-world use.

For temperature, the contactless approach benefits from the wrist support that improves repeatability through stabilized distance and posture. However, wrist skin temperature is a peripheral measurement influenced by perfusion and ambient conditions and does not necessarily coincide with core temperature trends. Moreover, IR-camera estimates depend on emissivity assumptions and environmental reflections. The presented results are therefore best interpreted as a validated *contactless skin temperature monitoring* capability under controlled geometry, rather than as a universal replacement for clinical core-temperature measurement.

The BP results must be interpreted considering dataset and reference constraints. In proprietary acquisition, BP reference values are oscillometric and intermittent, and they are paired with PPG segments acquired during the measurement interval; this makes the task inherently different from continuous beat-to-beat BP estimation. Additionally, population diversity remains limited (primarily healthy subjects), so generalization to pathological cohorts (arrhythmia, low perfusion, vascular disease) cannot be assumed. The PTT-based approach provides physiological interpretability but depends critically on robust ECG/PPG peak detection and on calibration choices; the PPG-only approach simplifies hardware but is sensitive to distribution shift across subjects and acquisition conditions.

The integration of the edge-BP workflow adds a further, deployment-specific limitation that is often overlooked in cuffless BP literature: floating-point performance is not a sufficient predictor of embedded deployability. Reviews on cuffless BP and edge-AI healthcare systems emphasize that end-to-end microcontroller validation is still relatively rare and that deployment constraints and hardware-aware optimization remain open challenges[35], [53].

In practice, full-integer quantization can alter error statistics in an architecture-dependent manner; therefore, model selection must consider quantization robustness as a primary criterion rather than a post-hoc implementation detail. The edge-oriented results support the view that accuracy, memory footprint, and numerical stability must be co-optimized for real-time wearable/surface-integrated use.

Finally, although the on-device STM32N6 validation demonstrates feasibility within strict memory and latency constraints, additional work is required for real-world translation. First, the embedded evaluation is performed on a public dataset with specific preprocessing and segmentation choices; broader clinical validation across diverse populations and prospective acquisition protocols remains necessary. Second, integrating preprocessing fully on-device (including artifact control and sensor handling) and characterizing energy consumption are essential steps for always-on operation. Third, acceptance thresholds commonly reported for BP devices should be interpreted within the context of standardized validation protocols rather than as universal guarantees across datasets and conditions[56].

4 Optical Fiber-Based Electronic Skin

Parts of this chapter are based on material previously published in “Edge AI Algorithm for FBG-Based E-Skin Touch Localization on Embedded Electronics”, E. Leogrande et al.[57], and “Electronic skin technologies: From hardware building blocks and tactile sensing to control algorithms and applications”, E. Leogrande et al.[58]. Compared with the published version, the present chapter reorganizes and extends the content to provide additional methodological details, broader discussion, and a clearer integration within the overall thesis framework.

Collaborative robotics is a major driver of the Industrial Revolution, promoting human-machine cooperation within shared workspaces[59]–[61]. This paradigm is rapidly expanding across application domains, including healthcare settings[60]. However, close physical proximity introduces new safety and reliability requirements, making robust human-machine interfaces a critical research and engineering challenge[61]. To address these needs, robots are increasingly equipped with multiple sensory modalities that replicate key aspects of human perception[15]. Among them, touch sensing is commonly achieved through flexible, thin, and lightweight artificial electronic skins (e-skins) that can be seamlessly integrated onto robotic platforms[16]. For safe and intuitive interaction, e-skins are expected to provide estimates of relevant physical parameters such as contact location and force, supporting robotic awareness of the surroundings and enabling prompt reactions during interaction[17], [26], [62], [63].

At the same time, the adoption of AI in Industry 4.0 has been accelerated by the synergy between AI and Big Data[64], with a growing impact on robotics and control strategies[65], [66]. Yet, AI models often require significant computation and handle large data volumes, which has traditionally favored cloud deployment[67]. Cloud-based solutions, however, face latency, bandwidth, security, and privacy limitations[68]. Since timing is a key constraint in collaborative manipulation and safety-critical interaction[17], [36], the trend toward edge AI is strengthening[28], [68]. Edge deployment mitigates distance-to-data issues but introduces power and memory constraints, frequently requiring model design choices and compression/pruning that may affect performance[28], [35], [69]. Nonetheless, MCU-class embedded systems offer low power consumption and small footprints, enabling real-time, on-site applications and fostering TinyML adoption[70].

Within this framework, this chapter presents an FBG (Fiber Bragg Grating)-based tactile e-skin and an edge AI pipeline for touch localization deployed on embedded electronics.

4.1 Background on Optical Fiber Sensing for Robotics

Robot skins are widely regarded as enabling technologies for safe collaboration, immersive teleoperation, and affective interaction in future collaborative robots[16]. In tactile sensing, the core requirement is to infer meaningful interaction variables (e.g., location and force) from distributed sensory signals, which is especially relevant in compliant or soft materials where deformation spreads spatially[17], [26], [62], [63].

Optical fiber sensing, specifically FBG technology, offers characteristics that are attractive for large-area tactile skins, including a favorable form factor and multiplexing capabilities that reduce wiring complexity while supporting multiple sensing points[63]. In FBGs, the Bragg wavelength intrinsically encodes sensitivity to deformation, depending on effective refractive index and grating pitch[71]. When strain is applied, the back-reflected spectrum peak shifts from its original center wavelength, providing a measurable signal proportional to deformation[72]. These properties motivate FBG-based e-skins as viable tactile layers for collaborative robotics, including deep-learning-based tactile perception approaches demonstrated in prior work[63].

4.2 System Design and Fabrication

The system design and fabrication of the proposed FBG-based e-skin were guided by two main requirements: (i) embedding a multiplexed optical sensing element into a compliant substrate to reliably convert indentations into measurable wavelength shifts, and (ii) defining an acquisition and embedded-processing architecture that supports both offline dataset generation and real-time touch localization on edge hardware.

4.2.1 Sensing Principle and Materials

The proposed tactile e-skin consists of an 8 mm-thick silicone substrate (Dragon Skin 10 Medium, Smooth-On Inc., USA) embedding an optical fiber within its middle layer[63]. The fiber includes 21 FBG units, each 8 mm long, with Bragg wavelengths spanning 1520-1580 nm (FBGS International NV, Belgium). The substrate is produced by mould casting: a first Dragon Skin layer is poured into a mould that defines a dedicated channel for the optical fiber; the FBG-bearing fiber is then laid into this channel and a closing silicone layer is cast on top, so that the fiber is fully encapsulated in the medial plane of the substrate. Placing the fiber at mid-thickness protects it mechanically and ensures that surface contact loads are transferred to the gratings, while the embedding depth governs the trade-off between local sensitivity and the spatial extent of the per-grating sensing regions.

Each FBG is a periodic modulation of the refractive index photo-inscribed along the fiber core, and it behaves as a wavelength-selective mirror: of the broadband light travelling in the fiber, only a narrow band centred on the Bragg wavelength is back-reflected, while the rest of the spectrum is transmitted. The reflected peak satisfies the Bragg condition $\lambda_B = 2n_{eff}\Lambda$, where n_{eff} is the effective refractive index of the guided core mode and Λ is the grating pitch, i.e. the spatial period of the refractive-index modulation. Because the

21 gratings are inscribed with distinct nominal pitches, their Bragg wavelengths are spectrally separated across the 1520-1580 nm band, so that each sensing point can be interrogated independently along a single fiber through wavelength-division multiplexing. Any mechanical perturbation that alters Λ or n_{eff} shifts λ_B accordingly, and this wavelength shift is the quantity measured. Under indentation, strain propagates through the silicone and is transferred to the embedded fiber, producing local shifts of the reflected peaks (Fig. 4.1). Since the e-skin is loaded only by indentation (the interaction presses the surface and never stretches it), the gratings are perturbed by a single, unilateral contact load rather than by an externally imposed tension–compression cycle. The transferred axial strain changes the grating pitch Λ, and thus the Bragg wavelength: a local elongation increases Λ and shifts λ_B toward longer wavelengths, whereas a shortening decreases Λ and shifts it toward shorter wavelengths. In the acquired data, the magnitude of the de-offset shifts grows consistently with indentation depth, while the sign of the shift varies with the grating position relative to the contact; a detailed strain characterization was beyond the scope of this work. The resulting position-dependent, multichannel time series encode the contact-induced deformation pattern across the embedded sensing line.

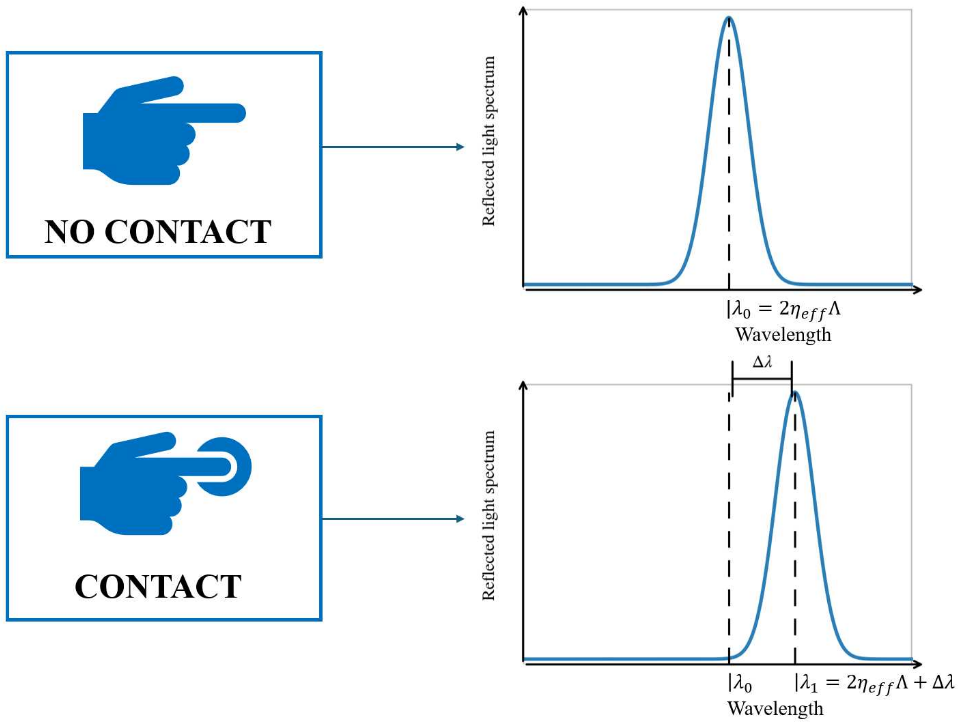

Figure 4.1. *FBG working principle: applied strain is encoded as a shift in the reflected*

spectrum peak wavelength.

4.2.2 Electronics and Acquisition Architecture

Data for training and deployment were collected through 2074 indentations over the e-skin surface, performed by a bimanual robotic platform (two anthropomorphic Racer 5L robots, Comau, Italy) using a force-controlled protocol. The FBG responses (time-varying wavelengths) were acquired with an optical interrogator (FBGS International NV, Belgium) at 100 Hz and stored in text files for analysis. Robotic control and acquisition routines were implemented in LabVIEW (National Instruments, USA).

For embedded real-time inference, the trained CNN was ported onto an STM32F767ZI MCU (32-bit Arm Cortex-M7) using X-CUBE-AI within STM32CubeMX. Network parameters were compressed to fit MCU memory constraints and validated on both desktop and target hardware. For streaming and visualization, Ethernet communication relied on the LwIP stack with UDP (LAN8742 PHY), and an RTOS supported host-MCU data exchange. A LabVIEW GUI transmitted FBG data from the interrogator to the MCU and visualized target/predicted points on a 3D skin model.

The overall acquisition and embedded inference workflow, including UDP streaming between the host interface and the target MCU and the feedback of contact-location predictions, is summarized in Fig. 4.2.

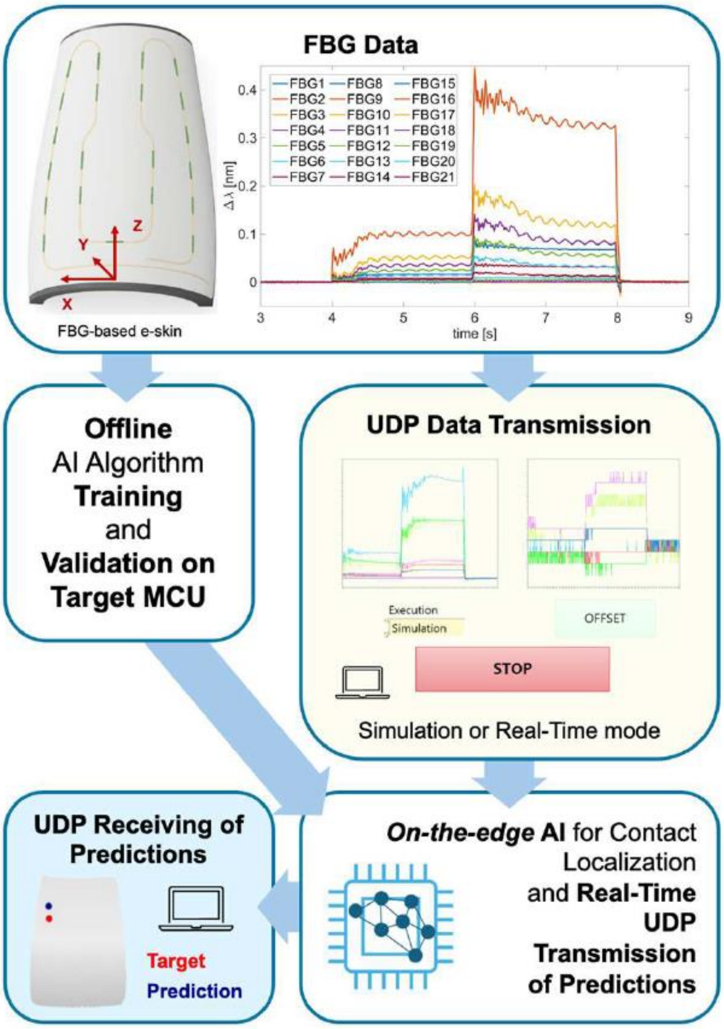

Figure 4.2. *End-to-end architecture of the proposed FBG-based e-skin system: FBG wavelength-shift data acquisition, offline CNN training/validation, UDP-based streaming to the target MCU for on-the-edge contact localization, and UDP return of predicted contact coordinates to the host GUI (simulation or real-time mode).*

4.3 Contact Detection and Interaction Sensing

The interaction task addressed here is touch localization, i.e., estimating the Cartesian coordinates (x, y, z) of indentation points from the multichannel FBG signals. This aligns with the broader requirement for e-skins to provide contact location (and potentially force) for safe, responsive collaborative interaction[17], [26], [62], [63].

Pre-processing and dataset construction. For each indentation, the wavelength signals were de-offset by subtracting the initial value, yielding Δλ time series. Signals were segmented into non-overlapping windows of 100 samples (1 s at 100 Hz), shuffled, and split into training (7259), validation (1555), and test (1556) windows.

CNN model. A CNN was trained offline to regress (x, y, z) from the Δλ input matrices. The architecture comprised 6 convolutional layers, followed by flattening and fully connected layers, and a final output layer with 3 linear neurons (Fig. 4.3). Training used batch size 32, up to 120 epochs, and early stopping with patience = 10; MSE served as loss, while MAE and RMSE were used for evaluation.

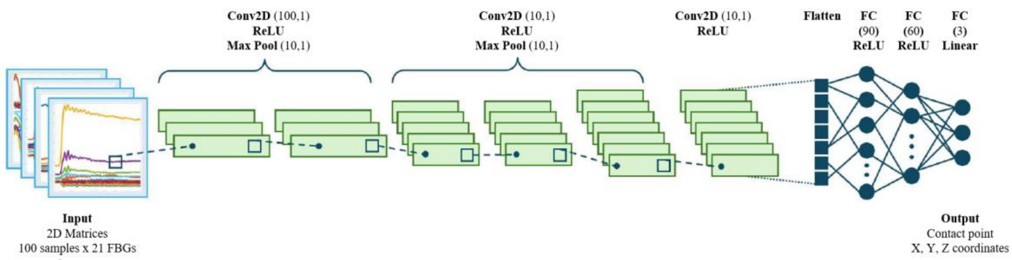

Figure 4.3. *Architecture of the implemented CNN.*

Spatial localization metric. In addition to coordinate-wise errors, localization was evaluated via the Euclidean distance between target and predicted points projected onto the skin XZ plane:

$$d = \sqrt{(x_p - x_t)^2 + (z_p - z_t)^2} \tag{7}$$

where (x_p, z_p) and (x_t, z_t) are predicted and target projected coordinates.

4.4 Experimental Validation and Results

The CNN was trained for 103 epochs to avoid divergence between training and validation losses. Test-set performance achieved MAE < 4 mm across Cartesian coordinates, as reported in Table 4.1.

Table 4.1. *Offline algorithm performance metrics.*

Set	MAE_x (mm)	MAE_y (mm)	MAE_z (mm)	RMSE_x (mm)	RMSE_y (mm)	RMSE_z (mm)
Training	1.89	1.38	2.71	4.23	2.25	5.43
Validation	2.25	1.60	3.94	4.53	2.54	6.83
Test	2.30	1.60	3.99	4.83	2.65	7.32

The Euclidean distance distribution on the XZ plane was generally low (mean 5 mm), with sparse localized hotspots exceeding 10 mm and reaching higher values in a limited number of locations (Fig. 4.4).

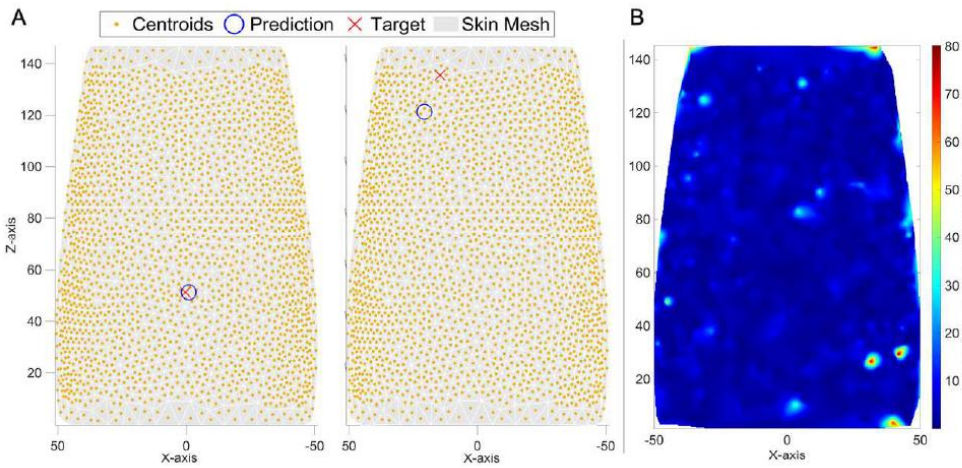

Figure 4.4. *Performance of the implemented AI algorithm. A) Two examples of offline predictions on test indentations. B) Distribution of the Euclidean distance values over the skin surface.*

Embedded deployment. For the on-the-edge implementation, CNN weights were compressed by a factor of 4 (761,488 vs 2,943,412 for the float model). The deployed model required 53.81 KiB RAM and 5,140,854 MACC, and it was validated on target on 10 random inputs with a total execution time of 12.22 ms. Comparing embedded outputs against the trained model reference yielded RMSE 0.36 mm and MAE 0.26 mm, indicating limited numerical deviation due to deployment.

An example of real-time touch localization obtained with the CNN deployed on the target MCU is shown in Fig. 4.5.

Figure 4.5. *Example of real-time localization of a touch event by the on-the-edge AI algorithm deployed on the MCU.*

4.5 Discussion and Conclusions

This chapter described an FBG-based tactile e-skin combined with a CNN for touch localization and its deployment on embedded electronics for edge AI inference. The approach is consistent with the need for e-skins to provide contact location information to support safe, responsive collaborative robotics[17], [26], [62], [63], and it builds on the established relevance of robot skins as interaction-enabling sensing layers[16].

From an architectural standpoint, moving inference from cloud to edge addresses latency, bandwidth, and privacy/security concerns[68], while meeting stringent timing constraints in human-robot collaborative manipulation[36]. At the same time, embedded deployment introduces memory/power limits that motivate model compression and optimization strategies, sometimes at the expense of performance[28], [35], [69]. In the reported results, deployment caused only a slight accuracy drop, while enabling real-time qualitative demonstrations on MCU hardware. Future work should focus on quantitative characterization of real-time performance and on systematic optimization/validation strategies on the target device, as also recommended by practitioner-oriented guidance on model compression workflows[34], [73]. Finally, intelligent robot skins combined with edge AI and

TinyML can provide real-time tactile perception directly on embedded hardware, supporting fast and reliable contact-aware behaviors in collaborative robots[33], [74].

5 Capacitive Electronic Skin for Proximity and 3D Object Localization

Parts of this chapter are based on material previously published in “Learning Contact Localization for Proximity-Tactile Robotic Skin from Real Single-Object Data and Synthetic Multi-Object Data from Data-driven Surrogate Model”, E. Leogrande et al. (submitted to Intelligent Service Robotics). Compared with the published version, the present chapter reorganizes and extends the content to provide additional methodological details, broader discussion, and a clearer integration within the overall thesis framework. This chapter is more extensive than the preceding ones, and the difference is intentional. Unlike the FBG-based skin of Chapter 4, where 21 multiplexed channels make touch localization a well-conditioned problem, the capacitive-pneumatic skin must resolve spatial information from only five channels. This sparsity makes the inverse problem severely under-constrained and motivates the more elaborate hardware-refinement and learning pipeline described here, which accounts for the greater length and level of detail of the chapter.

5.1 Background on Capacitive Sensing for Human-Robot Interaction

Tactile and proximity perception play a central role in human-robot interaction, especially in scenarios where robots are expected to operate safely in close physical proximity to people. In these contexts, sensing should not be limited to detecting collisions after they occur, but should also provide information about approaching objects, incipient contact, and interaction location over the robot surface[75], [76]. This requirement is particularly relevant in healthcare and assistive environments, where robotic systems may need to work near fragile users, support manipulation, or enable more natural physical interaction. For these reasons, surface-integrated sensing has emerged as a key enabling technology for safer and more informative physical interaction, allowing the robot body itself to become an active sensing interface rather than a passive mechanical structure[77], [78].

Among the sensing solutions proposed for robotic skins, capacitive sensing is especially attractive because it is compatible with compliant and large-area implementations, requires relatively simple electronics, and can provide information not only under direct contact but also in proximity conditions[79]–[81]. In capacitive e-skins, the measured signal depends on the electric field distribution generated by the electrodes and on how this field is perturbed by the presence of a nearby object or by mechanical interaction with the surface. As a result, capacitive sensors can detect the approach of a target before contact and can encode spatial information through the different responses observed across multiple electrodes. Recent capacitive platforms, including self-capacitance e-skins, non-array tactile sensors, and flexible capacitance-tomography systems, further show how sparse electrical measurements can support increasingly rich forms of proximity perception, object localization, and object-state inference[80]–[82].

A major challenge arises when a limited number of sensing channels must encode interaction over a relatively large surface. In this case, the inverse mapping from electrical measurements to object states becomes under-constrained. Even in single-object regimes, the signal depends jointly on planar position, standoff distance, and interaction intensity. In multi-object conditions, the problem becomes even more ambiguous, because different spatial configurations may produce partially similar signal combinations, especially when the sensing layout is sparse. This makes tactile inference resemble a super-resolution problem, where a small number of channels must support the reconstruction of richer spatial events from low-dimensional observations[83]–[86]. This difficulty is also evident in recent multi-touch tactile systems, including soft capacitive e-skins and tomographic sensing approaches, where force, location, and object multiplicity must still be inferred through learning or inverse reconstruction rather than read out directly from the hardware[87]–[90].

For this reason, many recent systems combine capacitive information with complementary sensing cues. In particular, pressure-related measurements can improve robustness in near-contact and contact regimes, where deformation and penetration provide additional information on interaction intensity[91], [92]. More broadly, the literature on electronic skin has shown that combining electrical and mechanical sensing channels can enhance robustness and expand the range of measurable interaction states[91]. However, a single pressure-related signal is generally not sufficient to resolve spatial ambiguity on its own, especially in the presence of multiple objects. A more effective strategy is therefore to treat capacitive and pressure signals as complementary modalities within a joint sensing framework, where capacitive channels contribute most of the spatial information and the pressure cue provides additional support during direct interaction. This combined perspective is particularly relevant for soft or compliant robot skins, where interaction involves both field perturbation and mechanical deformation[91], [92].

In parallel with sensor development, recent work has increasingly explored simulation-assisted learning as a way to address the data bottleneck in tactile perception. This direction is especially appealing in sparse tactile sensing, where collecting large, fully annotated real datasets is difficult and expensive, particularly in multi-object conditions. Simulation-driven or simulation-assisted tactile learning has been studied in several forms, including synthetic tactile data generation, force-map prediction, physics-based simulation, and sim-to-real calibration strategies[93]–[99]. Prior studies on electrical tomography-based tactile sensors have shown that learned simulators and sim-to-real transfer can help bridge the gap between controllable physical models and real tactile measurements[94]–[96], while more recent work has extended this idea toward zero-shot tactile data generation and differentiable tactile simulators for contact-rich interaction[93]. These contributions do not solve the multi-object problem directly, but they strongly motivate the use of simulation-assisted pipelines when tactile supervision is expensive or incomplete.

A further aspect that becomes relevant in multi-object perception is the temporal structure of interaction. Object events do not necessarily occur simultaneously; in many realistic situations they appear with short but non-zero onset delays. This point is not only algorithmically relevant, but also biologically plausible, since tactile discrimination has been shown to depend on the temporal features of stimulation[100]. From this perspective, the challenge is not merely to estimate object position from sparse measurements, but also to exploit the temporal organization of the signals when multiple objects interact or appear sequentially.

Starting from this background, the work presented in this chapter focuses on a capacitive-pneumatic electronic skin designed for proximity and 3D object localization over a compliant surface, together with a learning framework that infers object position from sparse multimodal tactile signals. The chapter does not treat capacitive sensing as an isolated algorithmic problem, but as the result of a co-design process involving electrode geometry, grounding strategy, signal acquisition, robotic data collection, synthetic data generation, and data-driven inference. In this perspective, robust hardware behavior is a prerequisite for meaningful learning, while simulation-assisted and temporally aware learning methods are necessary to extract usable spatial information from signals that are inherently indirect, sparse, and depth dependent. Beyond its role as a safety layer for collision avoidance, the same skin can serve as an active tactile interface when integrated on a collaborative robot, enabling richer forms of physical human-robot interaction. Compared with prior capacitive and proximity skins[79]–[81], the framework presented in this chapter targets joint 3D localization across proximity and contact regimes with only four capacitive channels, rather than focusing on contact detection or single-electrode characterization. With respect to multi-touch, tomographic, and tactile super-resolution approaches[85]–[90], [93], it does not rely on dense electrode arrays or purely synthetic supervision but builds on a data-driven surrogate model trained on real single-object measurements, extended to multi-object inference through procedural mixing and paired-sample domain adaptation.

5.2 System Architecture, Electronics, and Hardware Refinement

The proposed sensing platform was developed as a capacitive-pneumatic electronic skin intended to detect both proximity and object interactions over a relatively large surface while maintaining a limited number of acquisition channels. From a system perspective, the design combines three tightly coupled elements: a soft sensing surface, a sparse capacitive front-end able to capture spatially distributed field perturbations, and an auxiliary pressure-related measurement that becomes informative during direct interaction. Rather than treating signal processing as an isolated downstream step, the overall architecture was conceived as a co-designed sensing stack in which mechanical layout, electrode topology, grounding, and acquisition electronics directly influence the quality and interpretability of the data used for learning. This point is especially important in sparse e-skins, where the inference problem is already under-constrained and poor hardware stability can further amplify ambiguity.

The physical implementation is built around a compliant TPU layer forming a 16 × 16 cm skin, under which the sensing elements are arranged. The capacitive subsystem relies on four independent channels provided by a Texas Instruments FDC2214EVM, while the contact-related cue is acquired through a Honeywell ABP2DANT001BA2A3XX pressure sensor. This combination of capacitive proximity electrodes and a pneumatic pressure cue follows an approach explored in prior work on integrated proximity-tactile sensing[92]. Both sensing modules are interfaced to a Teensy 4.1 microcontroller through I²C communication and streamed to a host computer for synchronization, preprocessing, window construction, and model training (Fig. 5.1). In this configuration, the hardware does not merely provide raw measurements, but establishes signal conditions under which proximity, indentation depth, and planar location can later be inferred from the combined five-channel representation.

The capacitive and pressure front-ends rely on commercial modules (FDC2214EVM, Honeywell ABP2, Teensy 4.1), while the sensing surface, the four-electrode array on its TPU substrate, together with the dedicated grounding and shielding solution of Section 5.2.3, is custom-designed.

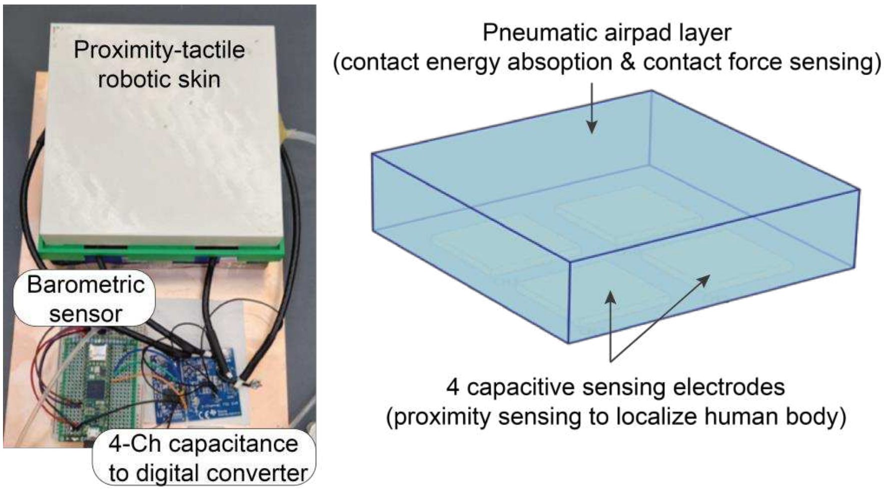

Figure 5.1. *E-skin Setup.*

5.2.1 Electrode Design and Sensing Topology

The capacitive sensing layout adopts a four-electrode topology arranged in a 2 × 2 configuration. Each electrode has dimensions of 5 × 5 cm, with 2 cm inter-electrode gaps and 2 cm margins from the outer borders of the 16 × 16 cm skin (Fig. 5.2).

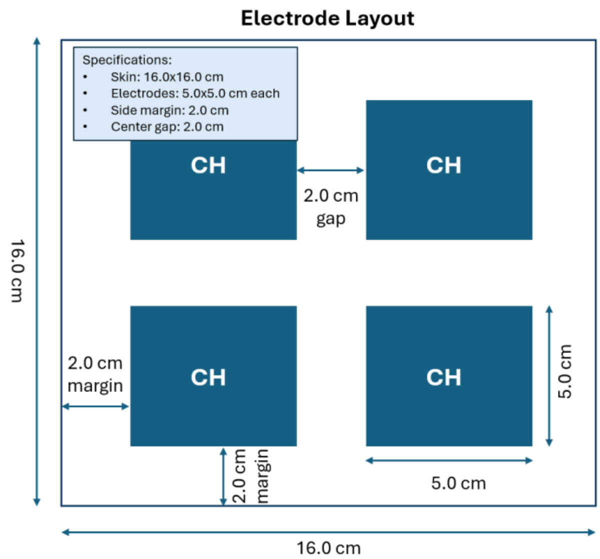

Figure 5.2. *E-skin Electrode Layout.*

This geometry provides a compromise between coverage area and channel count: the active surface is sufficiently large to support distributed interaction experiments, while the number of sensing channels remains low enough to keep the electronic architecture simple and scalable. At the same time, this sparse arrangement makes the localization problem intrinsically indirect, because the object state must be inferred from overlapping fringing-field responses rather than from densely sampled taxels.

From an operating viewpoint, each electrode senses changes in the surrounding electric field caused by the approach of an object or by deformation during contact. The resulting measurement pattern depends on the relative position of the indenter with respect to the four pads, so that localization information is encoded in the differential response across channels rather than in any single signal alone. For each channel, the sensed quantity is the baseline-subtracted capacitance $\Delta C = C - C_0$, where C_0 is the resting value measured in the absence of interaction, expressed in the raw output counts of the FDC2214.

Static characterization confirmed this expected behavior: when the indenter is placed above the center of a given electrode, that channel exhibits the dominant ΔC variation, while adjacent channels show weaker but still structured responses; conversely, positions located between two pads produce more balanced signal pairs, and the central region of the skin yields approximately symmetric responses across all four electrodes (Fig. 5.3). These observations support the interpretation of the array as a sparse spatial sensor in which each pad defines a localized but overlapping sensitivity region.

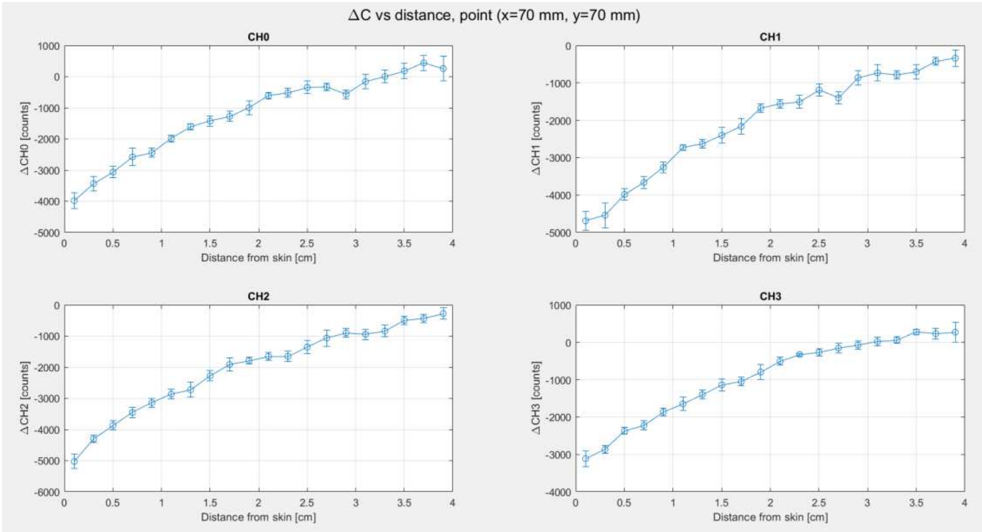

Figure 5.3. *Per-channel baseline-subtracted capacitance ΔC (in raw FDC2214 counts) as a function of the object distance from the skin surface, for an object approaching the center of the e-skin (x = 70 mm, y = 70 mm). Each panel corresponds to one of the four electrodes (CH0–CH3).*

The pressure sensor complements this topology by providing a contact-related cue that is mechanically coupled to the deformable layer. Its role is not to resolve planar position independently, but to enrich the sensing space in regimes where indentation and mechanical loading contribute meaningful information. In practice, the complete tactile sample is represented at each time step by a five-dimensional vector composed of the four capacitive channels and one pressure-related channel. This joint representation is especially useful because capacitive signals dominate spatial discrimination in proximity and early contact, whereas the pressure cue becomes more informative when penetration increases and the interaction enters a more clearly mechanical regime.

5.2.2 Embedded Acquisition Chain

The embedded acquisition chain is centered on the FDC2214EVM, which provides four independent capacitance-to-digital conversion channels connected to the electrode array. The device supports an input capacitance range of 0.01-250 pF, 28-bit digital output resolution, and configurable conversion rates between 10 and 400 Hz per channel, with a typical noise floor around 0.3 fF. In the present system, the board operates at 3.3 V and is used as the front-end for synchronous acquisition of the four capacitive measurements associated with the skin. This choice is appropriate for a prototype because it offers high sensitivity and multi-channel acquisition while remaining relatively straightforward to integrate with external controllers and custom sensing surfaces.

The pressure-related measurement is acquired in parallel through the Honeywell ABP2 sensor, and both sensing sources are interfaced to a Teensy 4.1 microcontroller via I^2C. The Teensy acts as a lightweight embedded bridge between the sensing hardware and the host workstation, handling data readout and streaming while leaving the computationally heavier stages, such as preprocessing, windowing, synthetic generation, and learning, to the PC side. This architecture reflects a pragmatic division of functions: the embedded layer ensures deterministic acquisition and hardware interfacing, whereas the host side supports flexible experimentation with dataset construction and model development. Within the final dataset, the acquired stream is sampled at 50 Hz and later segmented into non-overlapping windows of 50 samples, corresponding to 1 s temporal segments used as the basic learning unit.

Although the final learning framework is data-driven, the acquisition chain was intentionally designed to preserve as much raw temporal structure as possible. The preprocessing reported is deliberately lightweight: baseline subtraction is applied to the capacitive channels using a far-field reference region, and an analogous baseline correction is used for the pressure-related signal. This design choice avoids excessively engineered feature extraction and keeps the representation closer to the sensor dynamics produced by the hardware. Consequently, the quality of the recorded signals depends strongly on the physical stability of the electronics and interconnections, which motivated a dedicated hardware refinement effort.

5.2.3 Hardware Improvements and Grounding Strategy

An important part of the work concerned the progressive refinement of the electrical setup, because early hardware configurations exhibited signal artifacts that could not be ignored in a sparse capacitive skin. The first prototype of the hardware had long, temporary wires that initially behaved as unintended extensions of the electrodes, increasing susceptibility to environmental electromagnetic interference and introducing unstable parasitic effects. In this condition, measured capacitance no longer reflects only physical interaction with the target, but also the behavior of the interconnections and local grounding environment. This is a critical issue in capacitive sensing, where even small unintended capacitances can distort the relation between object position and channel response.

A more specific problem identified during the experimental campaign was the “ground-shift phenomenon”. Mixed ground return paths and insufficiently controlled shielding caused the coaxial cables to stop behaving as a true reference, thereby producing drifting signals, channel-dependent artefacts, and even non-physical trends such as capacitance variations growing with distance. To mitigate these effects, the setup was progressively redesigned by introducing shielded cables, shortening cable lengths, directly soldering coaxial shields to the FDC2214 ground, and finally adding a large copper and FR-4 ground plate connected both to the board ground and to earth ground (Fig. 5.4). In equivalent-circuit terms, this strategy adds fixed parasitic paths that stabilize the local reference potential and prevent uncontrolled ground motion.

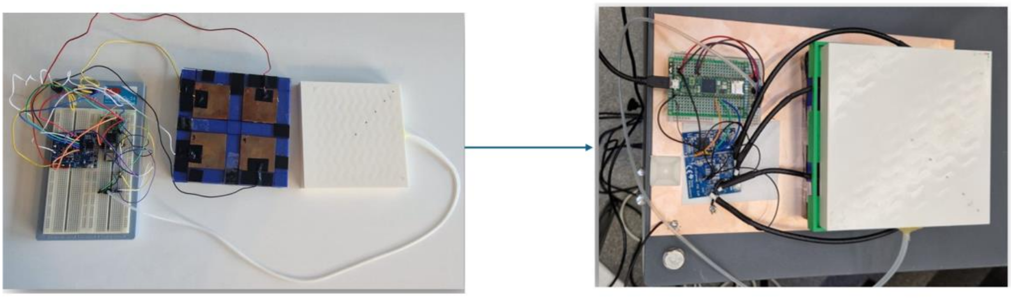

Figure 5.4. *Hardware Improvements.*

These hardware interventions were not secondary implementation details but enabling conditions for the rest of the study. The outcome is the removal of ground-shift effects, together with more stable and physically consistent capacitance signals, clearer monotonic ΔC-versus-distance behavior, and smoother 2D/3D spatial sensitivity maps (Fig. 5.5).

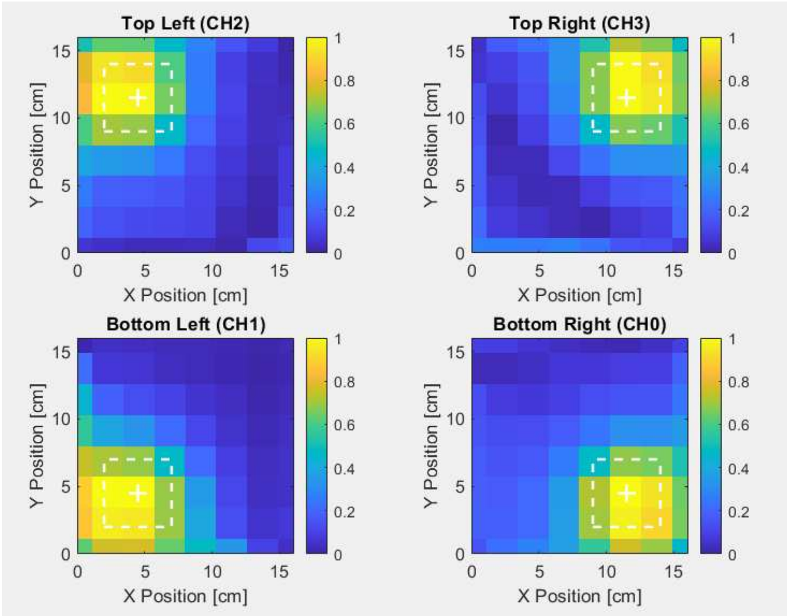

Figure 5.5. *2D maps showing the spatial sensitivity of each electrode during an 8×8 grid indentation test.*

This is an important methodological point: in a low-channel e-skin, the inverse learning problem is already ill-posed, so unreliable grounding and parasitic coupling do not simply add noise but can fundamentally corrupt the spatial structure that the model is supposed to learn. For this reason, the hardware refinement phase should be regarded as an essential part of the sensing methodology rather than as a purely practical adjustment.

5.3 Experimental Setup and Dataset Construction

The experimental methodology was designed to couple physically controlled interaction with a dataset structure suitable for both single-object learning and the later simulation-assisted extension to multi-object inference. In practice, the objective was not only to record sensor signals, but to do so under conditions where object position, depth, and interaction regime could be associated with reliable ground-truth labels. For this reason, the sensing skin was integrated into a robotic acquisition setup in which the motion of the indenter, the measured capacitive-pneumatic signals, and the corresponding spatial labels were synchronized throughout the data collection process. This choice is particularly important in sparse tactile sensing, where learning quality depends heavily on the consistency between the physical experiment and the label definition adopted during training and evaluation.

5.3.1 Static Characterization and Spatial Sensitivity Mapping

Before collecting the dynamic dataset used for learning, the hardware was first examined through static measurements aimed at verifying the spatial behavior of the sensor and the consistency of the four-electrode layout. This validation phase included controlled indentations or approaches at selected locations above the skin, with repeated measurements acquired over stable time intervals and then used to inspect the response of each channel as a function of position and distance. This preliminary step was essential to confirm that the electrical refinements described in the previous section had produced physically plausible sensing behavior before moving to large-scale robotic acquisition.

The resulting response patterns confirmed the expected topology of the array. When the indenter was positioned above the center of one electrode, that channel exhibited the largest capacitance variation, while neighboring channels showed weaker but still structured responses due to fringing-field overlap. Conversely, when the indenter was placed at intermediate positions between adjacent pads, the corresponding electrode pair produced comparable responses, and measurements acquired at the geometric center of the skin yielded approximately balanced amplitudes across all four channels. This behavior indicates that the sensor does not operate as a set of isolated taxels, but as a sparse distributed field-sensing surface in which localization information emerges from the relative combination of partially overlapping channel responses.

The same conclusion is supported by the 2D and 3D sensitivity maps obtained during the grid characterization. It is possible to localize peaks above the physical footprint of each electrode, followed by smooth spatial decay toward surrounding areas, effectively forming four distinct but overlapping sensitivity hills (Fig. 5.6).

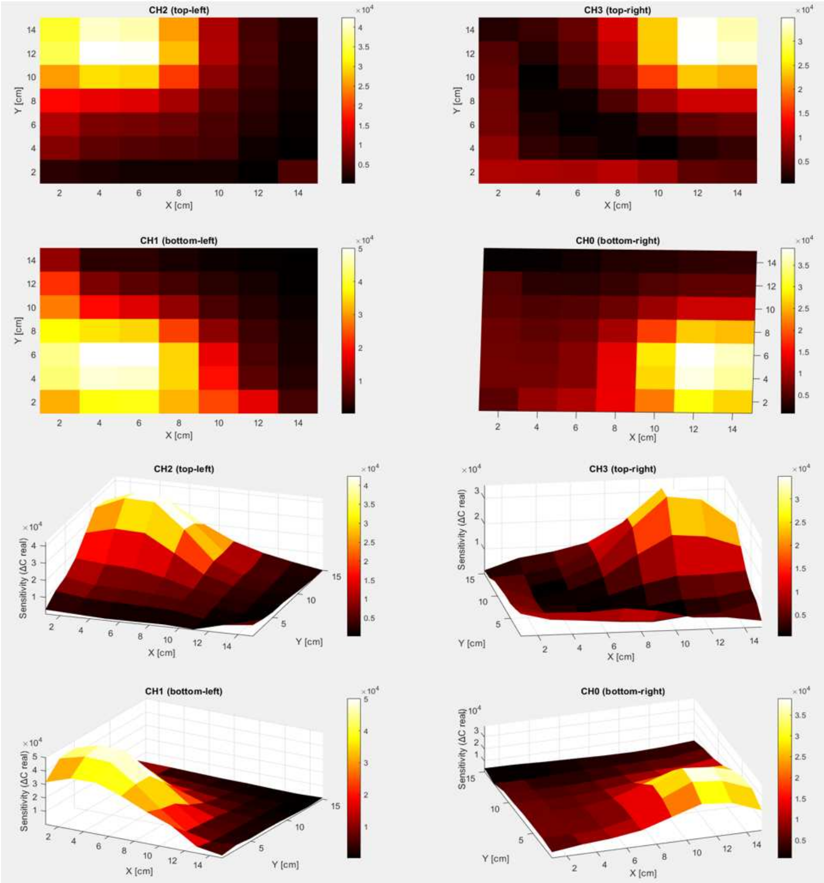

Figure 5.6. *Heatmap Grid Acquisition.*

These maps provide an important bridge between hardware and learning: they show that the measured signals retain a coherent spatial structure that can be exploited by data-driven models, while also revealing that the inverse problem remains indirect because no channel alone uniquely identifies the interaction point over the whole surface.

5.3.2 Robotic Acquisition Setup and Interaction Protocol

After static validation, the main dataset was collected using the UR5 robot, which applied controlled interactions over the skin while simultaneously providing the ground-truth end-effector pose used for labeling. A cylindrical indenter with 20 mm diameter was mounted on the robot end-effector to standardize the interaction geometry and reduce ambiguity associated with irregular or orientation-dependent contact areas (Fig. 5.7). Because of the indenter size, the effective usable workspace was restricted from the nominal 16 × 16 cm skin to a central region 14 × 14 cm, ensuring that the edge of the indenter does not reach the border of the TPU material throughout the acquisition domain.

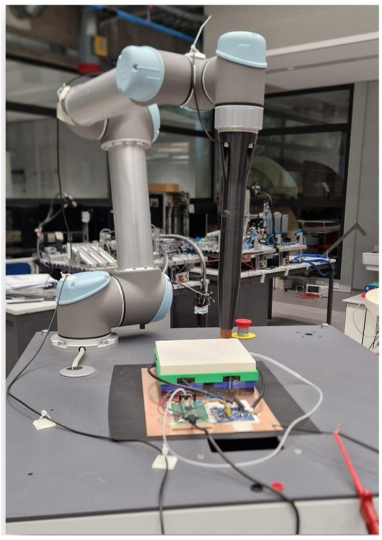

Figure 5.7. *Data Acquisition Setup.*

The acquisition protocol was designed to span both proximity and contact conditions within a unified coordinate convention. In the final dataset, depth is defined so that:

- $z = 0$ corresponds to surface contact
- $z < 0$ denotes penetration into the compliant layer down to -5mm
- $z > 0$ denotes increasing standoff distance in proximity

This convention is physically meaningful because it preserves continuity between the two regimes while making the transition from approach to contact explicit in the labels. Across the full campaign, the robot therefore sampled near-surface approaches, contact onset, and controlled indentation, allowing the recorded signals to reflect both capacitive field perturbation and pressure-related deformation.

A second strong point of the protocol is that the data were not collected through a single motion pattern. Instead, three trajectory families were used to expose the sensing system to complementary interaction structures. Grid trajectories provide structured sampling over the plane and are useful for spatial characterization and structured splitting. Parabolic trajectories introduce smooth curved motion with continuous transitions across proximity and contact. Random trajectories add pseudo-random coverage to reduce periodicity and to test robustness under less regular exploration paths. This choice reflects a methodological idea: a tactile model should not be trained only on one highly regular acquisition pattern and then expected to generalize automatically to different interaction dynamics.

5.3.3 Trajectory Design, Preprocessing, and Data Splitting

The dataset construction pipeline was kept intentionally simple at the preprocessing level to preserve the temporal structure of the raw measurements. At each step, the tactile observation consists of a five-channel vector formed by the four baseline-corrected capacitive signals and one baseline-corrected pressure-related signal. The data stream was sampled at 50 Hz and segmented into non-overlapping windows of 50 samples, corresponding to 1 s temporal segments that became the fundamental learning units. Each window was then associated with window-level labels describing mean x, mean y, mean z, and a discrete four-level force-related label. Importantly, however, the finalized study treats this force-related quantity only as an auxiliary variable for surrogate model conditioning and synthetic generation, not as a primary endpoint for performance evaluation.

Baseline correction was applied separately to the capacitive and pressure channels. The capacitive baseline was estimated from a far-field region and subtracted to obtain differential signals ΔC_0–ΔC_3, while the pressure cue was converted into a differential quantity Δp relative to its baseline level. This lightweight preprocessing is methodologically important: rather than imposing handcrafted features, it leaves most of the temporal and cross-channel structure available to the downstream learning models. The trade-off is that the dataset remains sensitive to acquisition consistency, which is precisely why the hardware refinement and robotic protocol discussed above are not peripheral details but necessary preconditions for robust learning.

The train/validation/test strategy was defined separately for each trajectory family, so that the data partitioning would remain consistent with the geometry and sampling logic of the acquisition protocol. Rather than relying on a uniform random split, the dataset was organized according to the structure of the underlying robot paths, with the aim of limiting leakage between subsets and obtaining a more meaningful assessment of generalization.

For the grid acquisition, samples were associated with their physical planar coordinates and indexed accordingly. The split was then built through a diagonal checkerboard assignment based on the sum of the grid indices modulo four, so that neighboring spatial locations were systematically separated across training, validation, and test subsets (Fig. 5.8). In practice, this prevents spatially adjacent points from being distributed across different splits, which would otherwise make the evaluation artificially optimistic because of the strong local similarity of the signals.

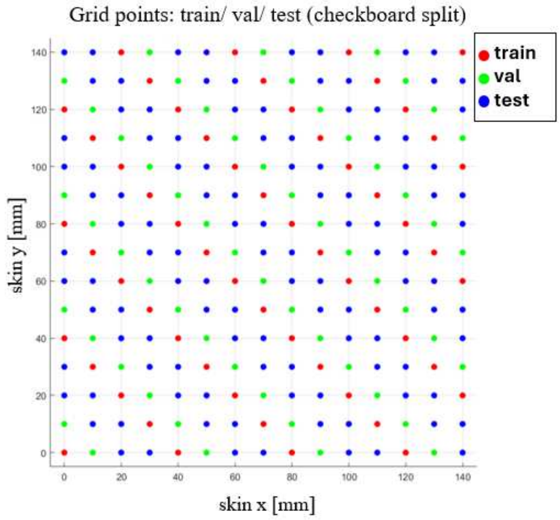

Figure 5.8. *Grid Trajectory.*

The parabolic trajectories were instead designed at three sampling densities, each corresponding to a different compromise between spatial resolution and path sparsity. The dense set consisted of 50 parabolas with points sampled every 3.0 mm, the moderate set comprised 36 parabolas with 5.0 mm spacing, and the coarse set included 20 parabolas with 7.0 mm spacing (Fig. 5.9). For this family, the dense and coarse files were split temporally within each recording, using the first 80% of samples for training and the remaining 20% for validation, while the moderate configuration was kept entirely as an unseen test condition. This choice allowed the evaluation to probe generalization not just to withheld samples, but to a trajectory configuration with an intermediate path density not observed during training.

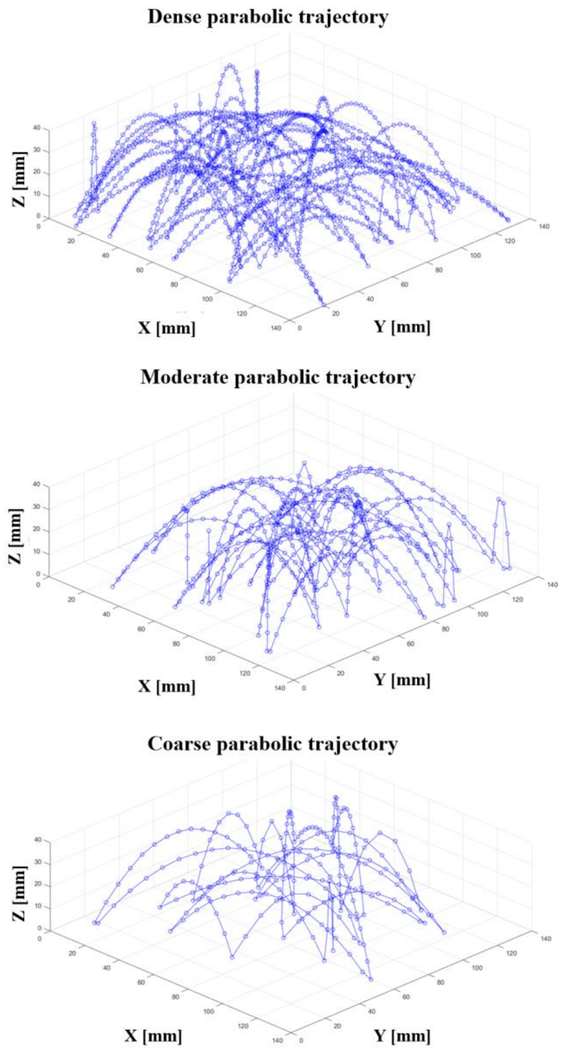

Figure 5.9. *Parabolic Trajectories.*

A similar rationale was adopted for the random trajectories, which were generated at three different densities as well. In this case, the dense subset contained 30 segments sampled every 10.0 mm, the moderate subset 20 segments sampled every 15.0 mm, and the coarse subset 10 segments sampled every 20.0 mm (Fig. 5.10). As with the parabolic family, the dense and coarse random files were split temporally into 80% training and 20% validation portions, whereas the moderate set was reserved entirely for test. This organization makes the test domain structurally different from the training one, thereby providing a more robust indication of how the model behaves on previously unseen motion patterns.

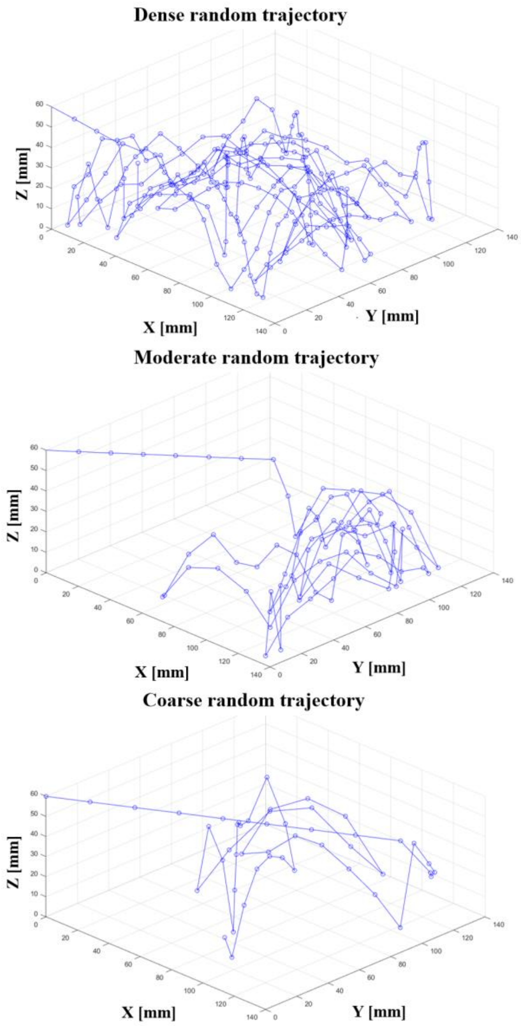

Figure 5.10. *Random Trajectories.*

After baseline correction, windowing, and split assignment, the resulting real single-object dataset comprised 7,225 training windows, 2,707 validation windows, and 3,868 test windows. Overall, this strategy ensured that the partitioning reflected the actual acquisition design and that performance was not inflated by overly permissive sample-level randomization.

5.4 Learning Framework for Object Localization

The learning framework adopted in this study was designed to address a central limitation of sparse capacitive-pneumatic skins: the availability of reliable real data is relatively high for single-object interactions but drops dramatically when moving to annotated multi-object conditions. The proposed approach starts with real single-object measurements and uses them as the basis for a simulation-assisted pipeline. In this way, the model is grounded in real sensor behavior at the single-object level, while synthetic generation and signal-domain adaptation are used to bootstrap multi-object inference under controlled assumptions. This strategy aims to extract richer spatial information from a low-channel-count sensing surface without requiring overly complex experimental annotation.

5.4.1 Single-Object Signal Representation and Labels

At the input level, each tactile sample is represented as a temporal window of 50 samples and 5 channels, corresponding to the four baseline-corrected capacitive signals and one pressure-related channel. With acquisition performed at 50 Hz, each window spans 1 s and preserves the short-term dynamics of the interaction without requiring handcrafted feature extraction (Fig. 5.11). This representation is deliberately simple: instead of reducing the signals to engineered summary descriptors, the framework operates directly on the baseline-corrected temporal traces, allowing the models to learn spatiotemporal patterns from the recorded sensor response itself.

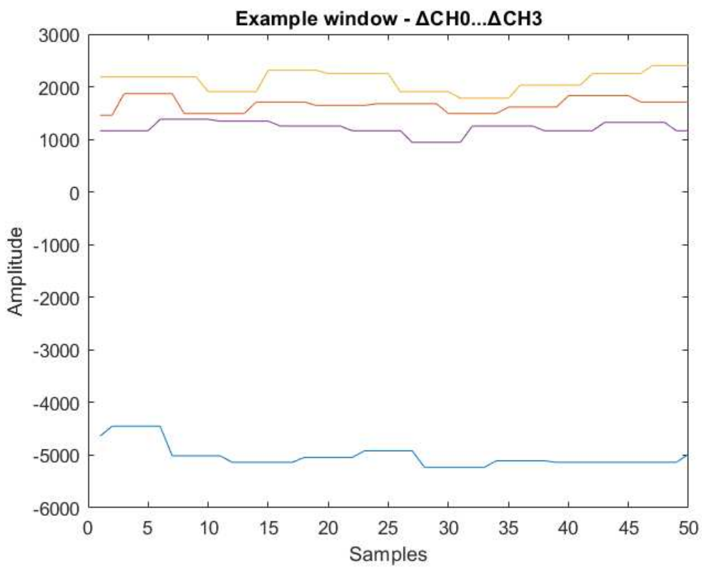

Figure 5.11. *Example window showing the acquisition of the 4 capacitive channels (pressure-related channel omitted for readability).*

Each real single-object window is paired with labels describing planar position, depth, and a discrete four-level force-related variable. However, a crucial methodological point is that the primary targets of the framework are the spatial quantities x, y, and z, whereas the force-related label is retained only as an auxiliary conditioning variable. This distinction matters because, in the recorded dataset, force level is strongly coupled to depth and contact regime; treating it as a primary, independently validated endpoint would therefore overstate what the data truly support. In the present formulation, the force-related label is useful because it enriches the description of the single-object state during simulation and synthetic generation, but the quantitative evaluation of the framework remains centered on localization and, in the temporal extension, object-number discrimination.

The single-object dataset also provides the reference domain from which all later stages inherit their physical meaning. Since the tactile windows are recorded under the same sensing hardware, the same coordinate convention, and the same preprocessing pipeline discussed in the previous sections, they define the only fully real domain available in this study. For this reason, single-object learning plays a dual role: it establishes a direct localization reference baseline on real measurements, and it provides the supervision required to learn how object-state labels map to sensor signals before attempting any multi-object inference. This is an important design choice, because it avoids treating the surrogate model as an abstract generator detached from the real acquisition process.

5.4.2 Simulation-Assisted Multi-Object Data Generation

The first stage of the proposed pipeline is a data-driven surrogate model trained on real single-object windows. Its purpose is to approximate the mapping from a recorded object-state label (x, y, z, F_{level}) to the corresponding normalized tactile window. The surrogate model is implemented as a multilayer perceptron (MLP) with two hidden dense layers of 64 and 128 neurons, followed by a final dense layer of size $T \cdot C$, whose output is reshaped to the window format $T \times C$. Conceptually, this network acts as a data-driven forward model: instead of solving the electrical and mechanical sensing equations explicitly, it learns from real single-object examples how the sparse capacitive-pneumatic skin responds over time to a given object state.

Once this single-object forward mapping is available, synthetic multi-object windows can be generated procedurally. The method samples one or two object-state labels under a minimum planar separation constraint, namely $\| (x_i, y_i) - (x_j, y_j) \|_2 \geq d_{\min}$, with $d_{\min} = 20\text{mm}$, simulates each object independently through the learned forward simulator, and then combines the resulting responses into a composite tactile window according to:

$$X_{mix} = clip(S(s_1) + \tau_\Delta(S(s_2)) + \in, -A, A)$$

where $S(\cdot)$ denotes the surrogate model, $s_i = (x_i, y_i, z_i, F_{level,i})$ is the state of object i, $\tau_\Delta(\cdot)$ is an optional temporal shift of Δsamples (with τ_0 corresponding to the non-sequential case), $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ is additive Gaussian noise in normalized space with $\sigma = 0.02$, and $A = 5.0$ is the clipping threshold. During this synthesis process, additional variations are introduced through small temporal shifts, additive noise, amplitude clipping, and depth-aware sampling. Specifically, the synthetic split is 100,000/12,000/5,000 windows for training, validation, and testing, respectively; temporal shifts are applied up to ± 2 samples; additive noise is introduced with a standard deviation of 0.02 in the normalized space; and amplitude clipping is set to 5.0. Moreover, depth-aware sampling is adopted so that 70% of the sampled objects are drawn from $z \leq 20$mm, in order to enhance coverage in the near-surface region where capacitive coupling is strongest. These choices are intended to prevent implausible overlap, enrich the training distribution in the most informative depth regime, and expose the inverse model to a broader range of realistic variations without requiring real multi-object acquisition.

This synthesis strategy is effective, but it also rests on assumptions that must be acknowledged explicitly. The combination of independently simulated objects does not fully reproduce all possible non-linear interactions of the real TPU-capacitive-pressure stack, especially under simultaneous or strongly overlapping deformations. A skeptical reading would therefore say that the framework is not learning "real multi-object physics" in a strict sense, but rather a physically constrained approximation anchored to real single-object behavior. The strength of the method is not that it eliminates the sim-to-real problem, but that it creates a practical route for generating structured multi-object supervision when fully annotated real multi-object data are unavailable.

5.4.3 Domain Adaptation and Permutation-Invariant Object Localization

Because synthetic mixtures generated by the data-driven surrogate model may still differ from real measurements due to forward-model bias and sensor/conditioning artifacts, the pipeline introduces a second stage to reduce the signal-domain gap between simulated and real tactile windows. A domain adaptation network is formulated as a residual 1D temporal convolutional network, defined as $X_{\text{ref}} = X_{\text{mix}} + g_\phi(X_{\text{mix}})$, where g_ϕ is a shallow temporal convolutional network. It is trained using paired single-object supervision: for each real recorded window X, the corresponding label is passed through the data-driven surrogate model to generate $\hat{X} = S_\theta(p)$, and the transfer model is optimized by minimizing the mean squared error between $T_\phi(\hat{X})$ and X. This is a key design decision, because it allows signal-domain alignment to be learned from data that are available, namely paired real single-object examples, without requiring any paired real multi-object data.

The final inverse stage is a permutation-invariant 1D CNN, that estimates up to two unordered objects from each tactile window. Specifically, the predictor outputs $K_{\max}$ object hypotheses in the form $\hat{Y} = P_\psi(X_{\text{ref}}) \in \mathbb{R}^{K_{\max} \times 4}$, where each of the K_{max} rows contain the predicted object state (x, y, z, F_{level}), and validity masks are used during training to handle windows with fewer than K_{max} objects. The model is trained using a permutation-invariant assignment loss that handles the unordered nature of the object targets. Multi-object localization is naturally a set prediction problem rather than a fixed-order regression task: if two objects are present, their physical identity does not depend on whether one is labeled "first" or "second." For this reason, the framework outputs up to $K_{\max} = 2$ objects hypotheses together with validity masks, and training uses a permutation-invariant assignment loss. During evaluation, predicted and true objects are matched through Hungarian assignment before computing the localization errors. This choice avoids artificial penalties due to label ordering and makes the evaluation better aligned with the physical nature of the problem.

Localization performance is reported through MAE, RMSE, and SD for x, y, and z. Performance is analyzed not only globally, but also by depth regime and by z-bins. This stratified evaluation is important because sparse capacitive sensing becomes more ill-posed as standoff distance increases and capacitive coupling weakens. A model that appears acceptable under a single averaged score may in fact fail systematically in far-field regions; by organizing the analysis across depth ranges, the framework makes this degradation explicit and provides a more physically interpretable characterization of performance.

5.4.4 Temporal Extension for Sequential Multi-Object Events

The original synthetic generation process produces non-sequential mixtures, where multiple objects coexist within the same window without explicit temporal ordering. However, this assumption is not always representative of real interaction, where objects events often occur with short but non-zero onset delays. To investigate whether temporal encoding can improve inference in such cases, the framework was extended with a sequential synthesis mode in which, for two-object samples, the first object appears earlier and the second object appears later within the 50-sample window, with configurable inter-object delay and rise envelope. This extension preserves the same depth convention used throughout the study, so that deeper contact still corresponds to lower z, and temporal effects are introduced without changing the physical interpretation of the labels. Based on this extended protocol, four localization model families were compared. The baseline model is a CNN trained on non-sequential synthetic windows. A temporal CNN adopts the same CNN predictor but is trained on sequential synthetic windows. A more expressive "*seq_gru*" model combines a CNN front-end with a BiGRU/GRU temporal encoder and is likewise trained on sequential windows. To improve robustness across temporal regimes, additional mixed-regime GRU variants were trained on sequential/non-sequential mixtures with ratios 45/55, 50/50, 55/45, and 65/35. For all full localization runs, synthetic training used 100,000 training samples and 12,000 validation samples, batch size 64, Adam optimizer with learning rate 5×10^{-4}, 60 epochs, early stopping, and learning-rate reduction. Common synthesis parameters were $k \in \{1,2\}$, minimum contact distance 20 mm, additive noise 0.02, time-shift augmentation of ± 2 samples, and z-biased sampling with 70% of sampled objects drawn from $z \leq 20$mm and 30% from $z > 20$mm.

The same temporal extension was also used to define an explicit object-number discrimination task, in which the model predicts whether a window contains one or two objects. Two binary classifiers with target $N_{\text{contacts}} \in \{1,2\}$ were trained on the same synthetic pipeline: "*baseline_count*", a CNN trained on non-sequential windows, and "*seq_gru_count*", a CNN+GRU model trained on sequential windows. These object-count models were trained with 80,000 synthetic training samples and 10,000 validation samples, batch size 64, learning rate 5×10^{-4}, 40 epochs, and early stopping. Evaluation followed a like-for-like fixed-test protocol, with identical random seed and identical fixed test set per domain across models, and permutation-invariant matching between predicted and ground-truth object sets. Two localization domains were considered, namely a fixed sequential test and a fixed non-sequential test.

For localization, the reported metrics were MAE_x, MAE_y, and MAE_z over the full z range, together with stratified metrics for $z_{\text{true}} \leq 20$mm and $z_{\text{true}} > 20$mm.

For the object-number analysis, fixed sequential test sets were further divided into delay bins (0-2, 3-6, 7-12, 13-20 samples) and strict near-overlap bins (0-0, 1-1, 2-2 samples), with reporting of count accuracy, recall for the two-object class, and specificity for the one-object class.

As in the rest of the framework, the discrete four-level force-related label was used only as an internal conditioning variable during simulator training and synthetic generation and was not evaluated as a primary output.

5.5 Experimental Results and Discussion

The experimental results should be interpreted as the outcome of a pipeline in which hardware stability, controlled acquisition, synthetic generation, and inverse learning are tightly coupled. A physically meaningful reading of the results must consider, first, whether the hardware produces spatially consistent signals; second, whether the domain adaptation stage effectively reduces the gap between simulated and real measurements; and third, whether the final localization models remain coherent with the sensing physics, especially across depth regimes where capacitive coupling progressively weakens. Under this perspective, the results support the overall feasibility of the proposed framework, while also making clear where its current limits remain.

A second interpretive lens concerns the epistemic status of each result. The validation logic of this chapter follows an explicit hierarchy: the real single-object test set is the only fully real, task-level localization benchmark, collected under the final hardware configuration and the same coordinate convention used throughout; the domain adaptation stage quantifies the residual sim-to-real gap on paired real measurements; the synthetic multi-object results demonstrate the internal consistency and plausibility of the pipeline, but are evaluated on generated mixtures rather than on a fully annotated real multi-object dataset; the sequential extension probes whether temporal encoding becomes informative when the signal formation regime contains explicit temporal structure. Real single-object results and synthetic multi-object results therefore carry different epistemic weight and are not intended as interchangeable benchmarks.

5.5.1 Hardware Validation and Sensitivity Analysis

Early setups suffered from long unshielded or insufficiently controlled connections, parasitic ground capacitances, and mixed return paths, producing drifting signals, channel-dependent artefacts, and even non-physical trends in which the measured capacitance changed inconsistently with distance.

The spatial validation carried out after the hardware refinements confirmed that the four-electrode array behaves coherently with its intended topology. When the indenter was positioned above the center of one pad, the corresponding channel consistently showed the dominant ΔC response, whereas adjacent electrodes produced weaker but structured signals due to fringing-field overlap. This behavior is important because it demonstrates that the sensor does not merely respond to contact in a binary fashion but preserves a structured spatial signature that can be exploited by data-driven models.

The 2D and 3D heatmaps acquired over the grid further reinforce this interpretation. Each electrode forms a localized sensitivity peak centered over its active area, with gradual decay toward neighboring regions, resulting in four overlapping sensitivity hills. From a methodological viewpoint, this is exactly the type of response needed for sparse tactile localization: enough spatial selectivity to distinguish regions of interaction, but also enough overlap to preserve continuity across the sensing surface. At the same time, the maps also reveal the intrinsic limitation of the system: because only four capacitive channels cover the full skin, the inverse problem remains under-constrained and must be solved through learned inference rather than direct geometric decoding.

5.5.2 Surrogate Model and Domain Adaptation Validation

The domain adaptation network is trained with paired single-object supervision, and its fidelity is evaluated on the full real single-object test set by comparing real windows against (i) the raw surrogate model outputs generated from the same recorded object-state labels, and (ii) the corresponding outputs after domain adaptation. In denormalized sensor units, the adaptation step reduces the global MAE from 1294.45 to 1198.92 (−7.4%) and the RMSE from 2352.35 to 2067.62 (−12.1%), while increasing Pearson correlation from 0.9754 to 0.9803. Similar gains are observed in normalized space, and per-channel relative errors decrease across all channels, with the largest improvement on the pressure-related channel ΔP, from 17.88% to 15.85% (−11.4%). These results confirm that the domain adaptation network measurably reduces the sim-to-real gap, while also making clear that a residual mismatch remains, particularly for slow baseline drifts and temporal conditioning effects, which is a useful reminder that surrogate-model fidelity is the foundation on which synthetic multi-object supervision rests (Fig. 5.12).

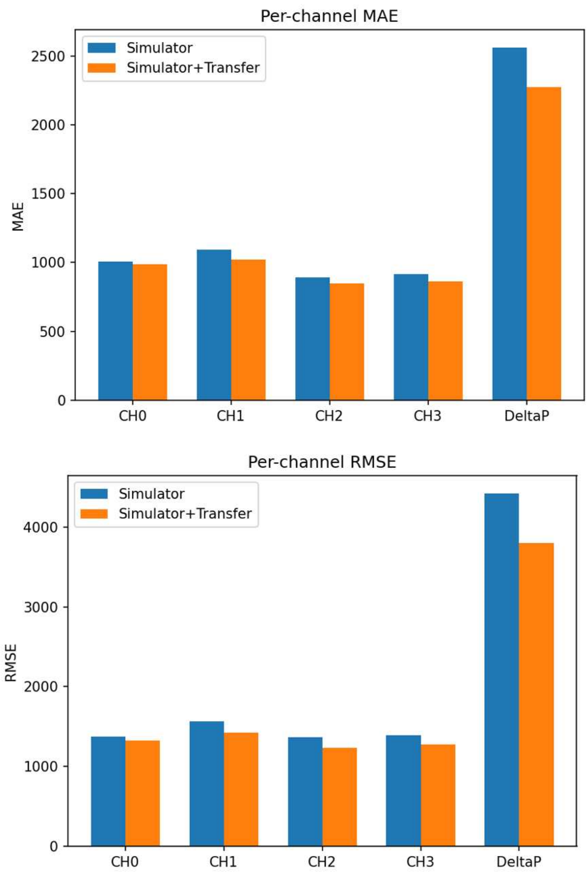

Figure 5.12. *Aggregate MAE/RMSE metrics.*

5.5.3 Single-Object Localization Reference Performance

The real single-object test set provides the most direct quantitative reference for the sensing system, because it is the only domain in which measured signals and ground-truth labels are both fully real and collected under the final hardware configuration. On this set, the system achieved global MAE values of 10.54 mm for x, 10.24 mm for y, and 2.64 mm for z, with corresponding RMSE values of 16.71 mm, 15.28 mm, and 4.23 mm, and SD values of 16.40 mm, 15.20 mm, and 4.18 mm, respectively (Table 5.1).

Table 5.1. *Global position error metrics (MAE, SD, RMSE) for real single-object predictions.*

	X [mm]	**Y [mm]**	**Z [mm]**
MAE	10.54	10.24	2.64
RMSE	16.71	15.28	4.23

SD	16.40	15.20	4.18

These results provide a consistent reference error scale for the same hardware, coordinate convention, and label definitions, and are therefore used as the baseline against which the later synthetic multi-object performance is contextualized.

In addition to the global metrics, the real single-object predictions were also analyzed as a function of the true *z* range (Table 5.2).

Table 5.2. *Depth-stratified errors (MAE, RMSE, SD) over true-z ranges for real single-object predictions.*

Z range	**[-5.0, 0.0] [mm]**	**[0.0, 5.0] [mm]**	**[5.0, 10.0] [mm]**	**[10.0, 20.0] [mm]**	**[20.0, 40.0] [mm]**
MAE	x: 6.38 y: 5.47 z: 0.92	x: 7.17 y: 7.23 z: 1.83	x: 7.40 y: 8.24 z: 2.61	x: 9.05 y: 9.03 z: 3.07	x: 15.14 y: 15.00 z: 3.77
RMSE	x: 8.73 y: 7.68 z: 1.34	x: 9.83 y: 9.77 z: 2.62	x: 10.44 y: 11.06 z: 3.50	x: 14.43 y: 12.55 z: 4.11	x: 22.40 y: 20.69 z: 5.58
SD	x: ±5.97 y: ±5.39 z: ±0.97	x: ±6.73 y: ±6.58 z: ±1.87	x: ±7.38 y: ±7.40 z: ±2.34	x: ±11.25 y: ±8.72 z: ±2.74	x: ±16.52 y: ±14.25 z: ±4.11

The same physically plausible trend later observed in the synthetic multi-object domain emerged in the real case as well, with progressively larger localization errors at increasing standoff distances, where capacitive coupling becomes weaker and the inverse problem is intrinsically more ambiguous (Fig. 5.13).

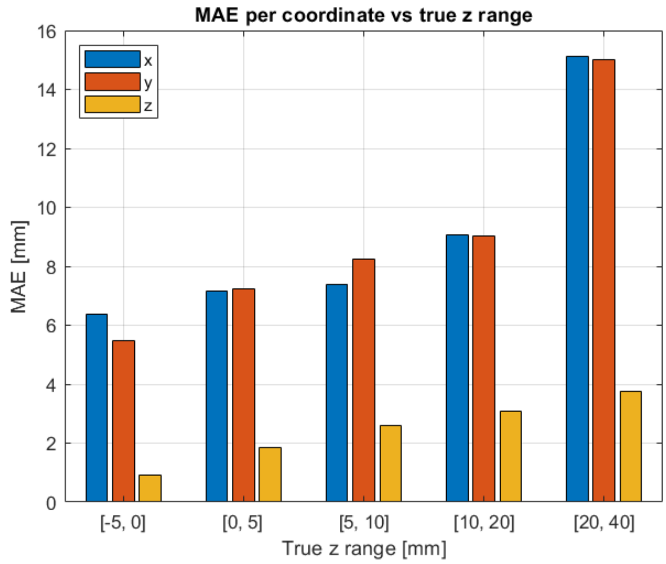

Figure 5.13. *MAE per coordinate vs true z range for real single-object data.*

This depth-dependent behavior is important because it confirms that even the real single-object benchmark already reflects the sensing physics of the sparse capacitive-pneumatic system, rather than an artificially simplified localization task.

These results are best understood as a realistic benchmark rather than as an upper bound to surpass mechanically. A superficial reading might suggest that single-object results should always be substantially better than multi-object ones. However, this expectation is not automatically justified. First, the single-object reference is measured on real data, whereas the multi-object evaluation is performed on synthetic mixtures generated under controlled constraints. Second, the sparse sensing layout makes even single-object localization an indirect problem, especially as the true z increases and capacitive coupling weakens. For this reason, the single-object reference does not represent a trivially easy setting but already reflects the ambiguity of sparse capacitive-pneumatic sensing over a relatively large area.

5.5.4 Synthetic Multi-Object Localization Results

On the synthetic multi-object test set after transfer, the pipeline achieved global MAE values of 8.47 mm for x, 8.79 mm for y, and 2.27 mm for z, with corresponding RMSE values of 14.35 mm, 14.75 mm, and 4.55 mm, and SD values of 14.35 mm, 14.74 mm, and 4.55 mm, respectively (Table 5.3).

Table 5.3. *Global position error metrics (MAE, SD, RMSE) for synthetic multi-object predictions.*

	X [mm]	Y [mm]	Z [mm]
MAE	8.47	8.79	2.27
RMSE	14.35	14.75	4.55
SD	14.35	14.74	4.55

This is not a direct like-for-like comparison with the real single-object case, but it shows that multi-object inference remains within a similar overall error scale when trained through the proposed sim-to-real procedure. Representative qualitative examples of true-versus-predicted multi-object configurations further illustrate that the model can recover plausible object sets under the adopted synthetic protocol (Fig. 5.14).

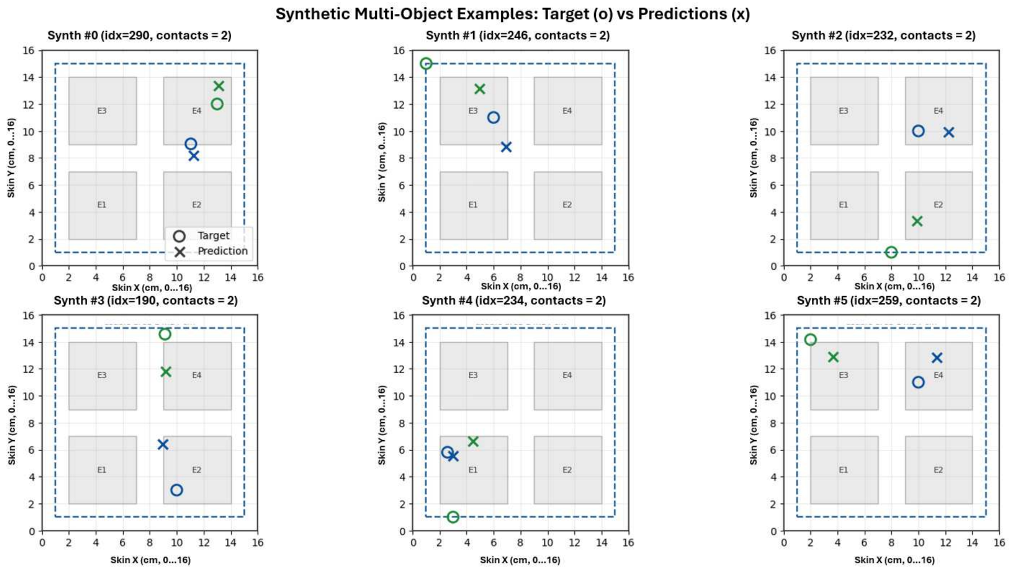

Figure 5.14. *Qualitative prediction examples (true vs predicted object locations) for representative synthetic multi-object samples.*

A more informative interpretation emerges from the depth-stratified analysis. Errors increase as true z increases, which is physically coherent with the sensing principle. In the contact regime $[-5,0]$mm, the MAE values are 4.58 mm, 5.73 mm, and 1.50 mm for x, y, and z, respectively. In contrast, in the farthest proximity range $[20,40]$mm, the MAE rises to 11.45 mm, 12.25 mm, and 2.94 mm. The same monotonic degradation is visible in RMSE and SD, with the most pronounced deterioration occurring in the farthest depth regime (Fig. 5.15).

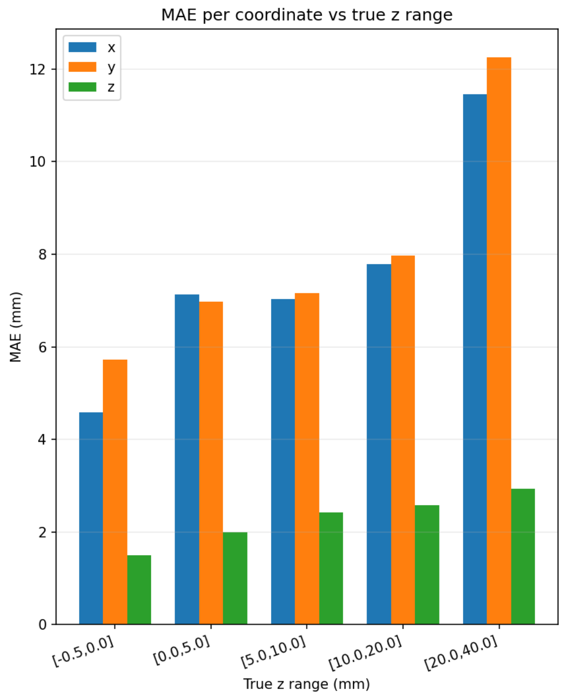

Figure 5.15. *MAE per coordinate versus true-z range, showing error growth with increasing distance from contact.*

Reporting performance by depth regimes and z-bins is therefore more informative than aggregate metrics alone, because it reveals where sparse-channel sensing becomes more ill-posed and how performance degrades across the joint capacitive-pressure domain (Table 5.4).

Table 5.4. *Depth-stratified errors (MAE, RMSE, SD) over true-z ranges for synthetic multi-object predictions.*

	[-5.0, 0.0] [mm]	**[0.0, 5.0] [mm]**	**[5.0, 10.0] [mm]**	**[10.0, 20.0] [mm]**	**[20.0, 40.0] [mm]**
MAE	x: 4.58 y: 5.73 z: 1.50	x: 7.13 y: 6.98 z: 2.00	x: 7.03 y: 7.16 z: 2.43	x: 7.78 y: 7.79 z: 2.57	x: 11.45 y: 12.25 z: 2.94
RMSE	x: 6.90 y: 9.45 z: 5.16	x: 11.20 y: 10.86 z: 4.34	x: 11.07 y: 11.68 z: 4.40	x: 12.53 y: 12.81 z: 4.37	x: 18.68 y: 19.28 z: 5.20
SD	x: ±6.90 y: ±9.17 z: ±5.08	x: ±11.19 y: ±10.86 z: ± 4.18	x: ±11.07 y: ±11.67 z: ±4.32	x: ±12.53 y: ±12.80 z: ±4.36	x: ±18.67 y: ±19.27 z: ±5.02

This depth-dependent behavior suggests that the model errors track the actual informativeness of the hardware rather than behaving arbitrarily. The progressive degradation with distance matches the expected decay of capacitive coupling and supports the validity of the adopted evaluation protocol. It also justifies the choice of reporting stratified metrics rather than only global averages, since aggregate scores alone would hide the regimes in which sparse sensing becomes most ambiguous.

5.5.5 Sequential Multi-Object Localization and Object-Number Analysis

To assess whether temporal encoding improves multi-object localization when objects occur sequentially in time, the procedural synthesizer was extended with a sequential mode in which contact onset order, delay, and rise envelope are explicitly controlled. Under this fixed sequential test domain, the best model was the GRU-based temporal predictor *seq_gru*, which achieved MAE values of 4.600 mm for x, 4.827 mm for y, and 1.077 mm for z. Temporal CNN ranked second, with 5.479 mm, 6.163 mm, and 1.288 mm, whereas the baseline CNN trained on non-sequential data performed substantially worse, reaching 13.745 mm, 14.521 mm, and 4.664 mm. In relative terms, seq_gru reduces MAE over the non-sequential baseline by 66.5% on x, 66.8% on y, and 76.9% on z on this fixed sequential test. Mixed-regime GRU variants showed intermediate behavior, with *mixed65_gru* and *mixed50_gru* providing the strongest compromises in the sequential domain after *seq_gru* (Table 5.5).

Table 5.5. *Sequential test: like-for-like comparison of localization models on the same sequential test set.*

	X_all [mm]	Y_all [mm]	Z_all [mm]
seq_gru	4.600	4.827	1.077
seq_cnn	5.479	6.163	1.288
mixed65_gru	6.049	6.480	1.364
mixed50_gru	6.483	6.441	1.401
mixed55_gru	6.591	6.852	1.496
mixed45_gru	6.750	6.938	1.527
baseline	13.745	14.521	4.664

These results show that temporal encoding becomes genuinely advantageous when the test domain contains ordered temporal structure.

At the same time, the results argue against a simplistic interpretation in which recurrent models are universally superior. On the fixed non-sequential test domain, the best model remained the original baseline CNN, with 8.421 mm, 8.751 mm, and 2.237 mm MAE for x, y, and z, respectively. Sequentially trained models degraded substantially in this regime, with seq_gru reaching 14.643 mm, 14.981 mm, and 4.712 mm. Symmetrically, seq_gru's MAE here worsens by about 74% on x, 71% on y, and 111% on z relative to the baseline CNN, confirming that the sequential advantage does not transfer to a regime lacking temporal structure. Mixed-regime GRU variants substantially reduced this degradation and consistently outperformed the purely sequential models, although they remained below the baseline CNN (Table 5.6).

Table 5.6. *Fixed non-sequential test: like-for-like comparison of localization models on the same non-sequential test set.*

	X_all [mm]	Y_all [mm]	Z_all [mm]
baseline	8.421	8.751	2.237
mixed45_gru	8.938	8.969	2.222
mixed50_gru	9.185	9.107	2.402
mixed55_gru	9.267	9.440	2.511
mixed65_gru	9.274	9.684	2.488
seq_gru	14.643	14.981	4.712
seq_cnn	17.064	16.208	4.641

This confirms that temporal models do not automatically generalize to non-sequential mixtures and that performance depends strongly on the match between training regime and test-time signal formation process.

The fixed-domain comparison therefore reveals a clear regime dependence. In the sequential test domain, temporally encoded models achieve lower localization errors, with seq_gru providing the best overall performance. In contrast, in the non-sequential domain, the baseline CNN remains the top performer, while purely sequential models show a marked degradation. Mixed-regime GRU variants provide intermediate behavior, partially preserving the sequential advantage while reducing non-sequential performance loss (Fig. 5.16).

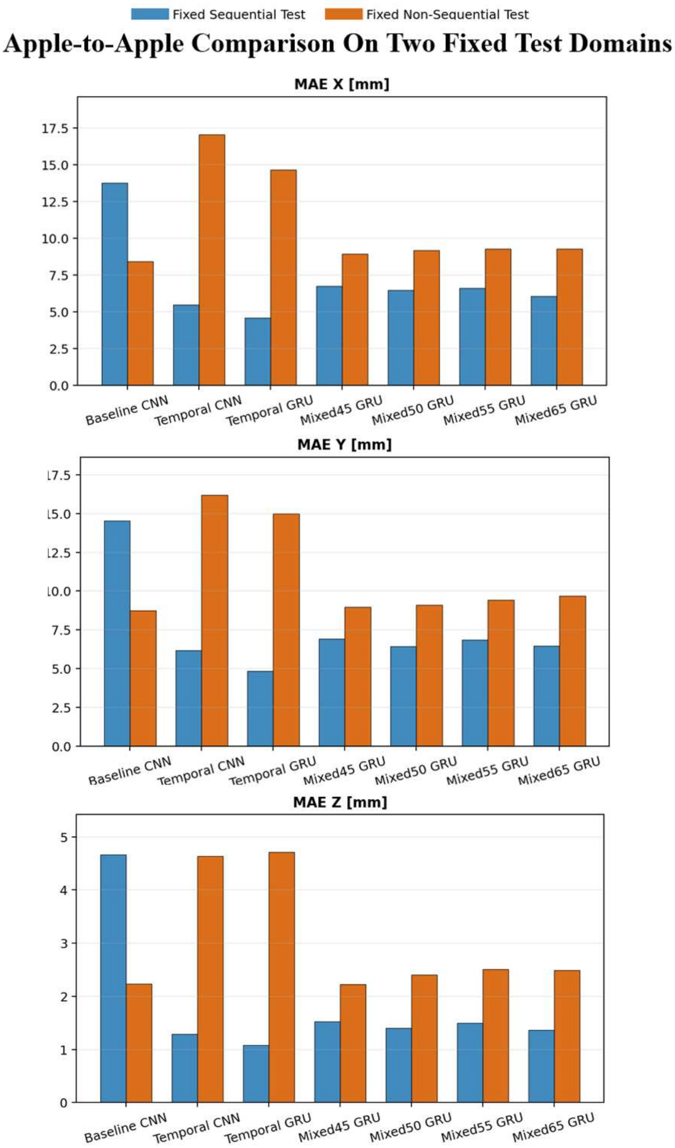

Figure 5.16. *Like-for-like comparison on fixed sequential and fixed non-sequential test sets. Metrics are MAE_x, MAE_y, and MAE_z.*

A complementary domain-gap analysis, computed as non-sequential minus sequential MAE for each model, further shows that mixed-regime training reduces regime shift relative to purely sequential training. Under equal domain weighting, *mixed50_gru* provides the best overall trade-off, whereas *mixed65_gru* becomes preferable when sequential-domain performance is prioritized (Fig. 5.17).

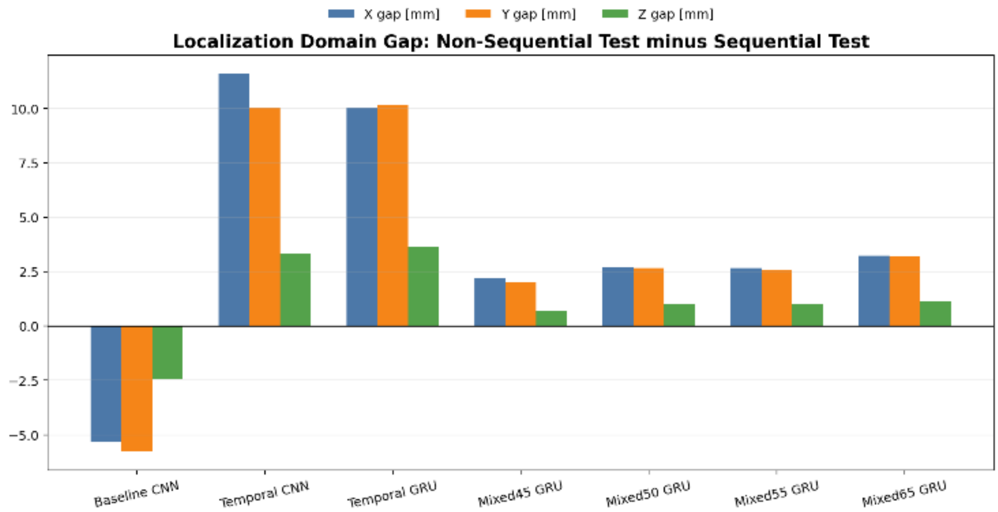

Figure 5.17. *Domain gap per model, computed as non-sequential minus sequential performance. Lower MAE gap indicates better cross-regime robustness.*

From a deployment perspective, these results translate into three practical guidelines: seq_gru is the preferred choice when the target scenario is known to preserve temporal structure between object onsets; mixed50_gru offers the most balanced behavior when the temporal regime cannot be assumed a priori, partially preserving the sequential advantage while limiting non-sequential degradation; mixed65_gru is preferable when sequential-domain accuracy is prioritized while a reduced domain gap is still desirable.

Additional delay-sweep analysis on the sequential domain, restricted to true two-object cases, confirmed that temporal separation is a key factor. At very small delays (0–2 samples), mixed models were more robust and *mixed50_gru* outperformed *seq_gru*. At larger delays (7–20 samples), *seq_gru* consistently became the best performer (Fig. 5.18).

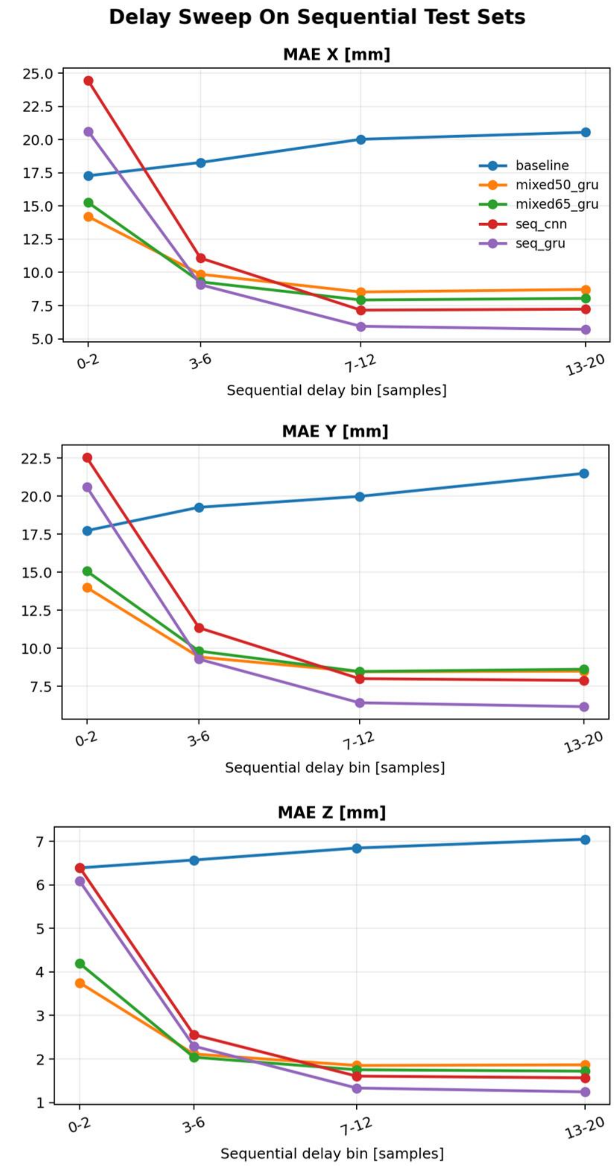

Figure 5.18. *Delay-sweep analysis in sequential two-object settings. Temporal models improve as inter-object delay increases, while mixed models remain more robust at very short delays.*

These results indicate that the benefit of temporal encoding strengthens as inter-object delay increases, while mixed-regime models remain advantageous in near-simultaneous settings (Table 5.7).

Table 5.7. *Delay-sweep summary for localization on the sequential domain (k=2 only).*

Delay bin	Best model	Note
0–2	mixed50_gru	more robust when objects are almost simultaneous
3–6	mixed65_gru, mixed50_gru, seq_gru	all strongly improved over baseline
7–12	seq_gru	stably best
13–20	seq_gru	stably best

A parallel conclusion emerges from the explicit object-number discrimination experiments.

Dedicated count models were trained on the same synthetic pipeline: *baseline_count*, a CNN trained on non-sequential windows, and *seq_gru_count*, a CNN+GRU model trained on sequential windows. On the fixed sequential test sets, *seq_gru_count* consistently outperformed *baseline_count* across delay bins, especially in recall for the two-object class as delay increased. For example, at delays 13–20 samples, *baseline_count* reached 85.06% accuracy and 70.54% recall for the two-object class, whereas *seq_gru_count* reached 100.00% for both metrics (Table 5.8).

Table 5.8. *Object number cue on fixed sequential test sets, evaluated across delay bins.*

Delay bin	Model	Accuracy	Recall (k=2)	Specificity (k=1)
0-2	baseline_count	98.66%	97.28%	100.00%
0-2	seq_gru_count	99.72%	99.43%	100.00%
3-6	baseline_count	94.90%	89.96%	99.96%
3-6	seq_gru_count	100.00%	100.00%	100.00%
7-12	baseline_count	91.04%	81.89%	99.96%
7-12	seq_gru_count	100.00%	100.00%	100.00%
13-20	baseline_count	85.06%	70.54%	100.00%
13-20	seq_gru_count	100.00%	100.00%	100.00%

A stricter near-overlap analysis showed that the gain is minimal at exact simultaneity, as expected when temporal ordering is absent. At delay 0–0 samples, both models are nearly equivalent, with 98.88% and 98.94% accuracy, respectively. However, already at delay 1–1 the GRU-based classifier reaches 99.96% accuracy compared with 99.06% for the baseline, and at delay 2–2 it reaches 100.00% compared with 98.24%. These results show that temporal encoding improves one-vs-two-object discrimination, especially when the inter-object delay is small but non-zero, whereas the gain becomes minimal at exact simultaneity, where temporal order is absent (Table 5.9). The effect is quantitatively substantial: at the 13–20 sample delay bin, recall for the two-object class rises from 70.54% with baseline_count to 100% with seq_gru_count, while at the 0–0 delay bin the two models are essentially equivalent. This is an important finding, because it indicates that temporal structure is not merely an architectural addition, but a genuinely informative cue when the signal formation process preserves short delay object separation (Fig. 5.19).

Table 5.9. *Object-number cue under strict near-overlap conditions.*

Delay bin	Model	Accuracy	Recall (k=2)	Specificity (k=1)
0–0	baseline_count	98.88%	97.77%	99.96%
0–0	seq_gru_count	98.94%	97.85%	100.00%
1–1	baseline_count	99.06%	98.16%	99.96%
1–1	seq_gru_count	99.96%	99.92%	100.00%
2–2	baseline_count	98.24%	96.47%	99.96%
2–2	seq_gru_count	100.00%	100.00%	100.00%

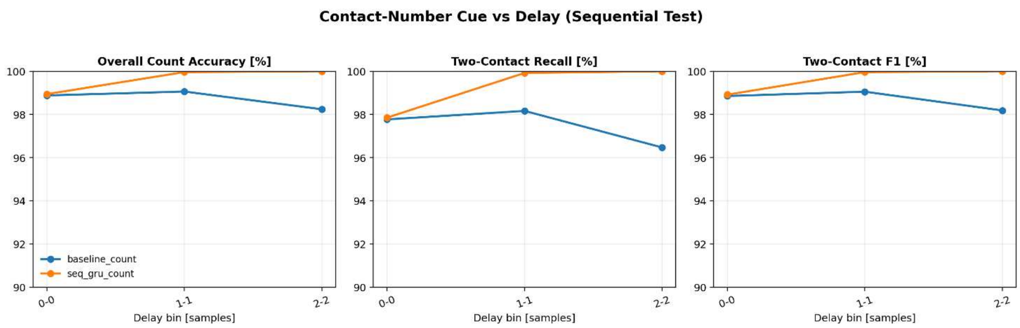

Figure 5.19. *Strict near-overlap object-number analysis on fixed sequential test sets (delay bins: 0-0, 1-1, 2-2 samples).*

5.5.6 Limitations and Discussion

The most important limitation is that the multi-object validation remains predominantly synthetic. Although the synthesis pipeline is grounded in real single-object measurements and further refined through paired single-object transfer, the final two-object evaluation is still based on generated mixtures rather than on a large fully annotated real multi-object dataset. This means that the framework demonstrates a plausible and well-structured bootstrap strategy, but not yet definitive validation under real multi-object sensing conditions.

A second limitation is that procedural mixing of independently simulated objects cannot capture all non-linear mechanical and electrical interactions that may arise in the real TPU stack. Simultaneous objects may interact through deformation coupling, pressure redistribution, and field perturbation effects that are only approximated by additive or weakly perturbed mixtures. For this reason, it would be misleading to describe the current synthesis process as a full physical simulator. A more accurate statement is that it provides a data-driven approximation constrained by real single-object behavior.

The framework is also restricted to at most two objects per window, and the force-related label is used only as an auxiliary conditioning variable rather than as a primary output. This is the correct methodological choice for the present study, because the force-related quantity is strongly coupled with contact depth in the recorded data and would not support a strong independent claim of force estimation. Quantitative evaluation is therefore focused on spatial localization and object-number discrimination.

The hardware refinement phase established physically consistent capacitive-pneumatic signals; the single-object results defined a realistic reference baseline; the domain adaptation network measurably improved surrogate model fidelity to real tactile signals; the synthetic multi-object pipeline achieved stable localization within a comparable error scale; the sequential extension showed that temporal encoding is beneficial when the signal-generation regime contains temporal order; and the mixed-regime GRU variants provided a practical compromise when robustness across temporal regimes is required. The framework therefore does not eliminate the intrinsic ambiguity of sparse e-skin sensing, but it provides a principled way to manage it using real single-object data, simulation-assisted synthesis, and regime-aware learning.

From a practical standpoint, purely temporal models are preferable when the deployment scenario is known to involve sequential object events, while mixed-regime temporal models represent the most robust choice when the temporal structure of inputs cannot be assumed in advance.

6 Comparative and System-Level Discussion

The three preceding chapters presented surface-integrated sensing systems that differ substantially in transduction principle, target application, and deployment architecture, yet share a common engineering logic: placing sensing on or near surfaces where interaction physically occurs and extracting actionable information through embedded electronics and data-driven inference. Chapter 3 addressed physiological monitoring through multi-wavelength photoplethysmography and infrared thermometry, with the sensing surface being the human body. Chapters 4 and 5 focused on robotic electronic skins, optical fiber-based and capacitive-pneumatic respectively, where the robot body itself becomes the sensing interface.

This chapter examines these three systems from a comparative and system-level perspective. The goal is not to rank them, since they address different interaction problems, but to extract cross-cutting observations about what makes surface-integrated sensing practically viable and where the main open challenges remain. The discussion is organized as follows: Section 6.1 distinguishes the different roles that surface sensing plays across the thesis; Section 6.2 provides a structured technical comparison across transduction, electronics, learning, and deployment; Section 6.3 distills shared methodological lessons; and Section 6.4 discusses implications for healthcare robotics deployment.

6.1 Different Roles of Surface-Integrated Sensing

A unifying observation across all three systems is that sensing is located where the relevant physical interaction takes place: on a body surface for physiological signals, and on the robot surface for contact, proximity, and approach events. However, the role of sensing differs fundamentally between the two application streams, and recognizing this difference is important for interpreting results and for understanding how the systems might eventually be integrated.

In the vital-sign monitoring stream (Chapter 3), the surface is the human body, and the primary objective is to extract physiological state information, specifically cardiovascular and thermal parameters, through non-invasive optical and infrared measurements. The robot or device, in this case, serves as a platform that hosts the sensing module. The sensing interaction is cooperative: the user places a finger on the sensor or positions a wrist within the thermal camera field of view. Motion artifacts and inter-subject variability are the dominant disturbances, but the contact geometry is relatively controlled, and the user is generally stationary during measurement. The output is a scalar or low-dimensional estimate (HR, SpO_2, temperature, SBP/DBP) that informs clinical or assistive decisions.

In the electronic-skin stream (Chapters 4 and 5), the surface is the robot body, and the objective is fundamentally different: to detect, localize, and characterize interactions that may be incidental, unplanned, or spatially distributed. Here the sensing must operate under conditions that are less cooperative, where contact can occur at any location, at varying intensity, and potentially involving multiple objects or users. The output is spatial and, in the multi-object case, combinatorial: the system must estimate where on its surface interaction occurs, at what distance or depth, and, ideally, how many distinct contact events are present. The two e-skin chapters differ in the granularity of sensing: the FBG-based system (Chapter 4) provides 21 multiplexed channels within a relatively small area, targeting sub-centimeter touch localization; the capacitive-pneumatic system (Chapter 5) covers a larger area with only 5 channels, requiring more aggressive inference to resolve spatial ambiguity.

Despite these differences, both streams confront the same fundamental challenge: extracting meaningful information from noisy, indirect surface measurements under deployment constraints that limit computation, latency, and available training data. In the physiological domain, the indirectness lies in the mapping from optical absorption to cardiovascular parameters; in the tactile domain, it lies in the mapping from sparse electrical or optical measurements to spatial interaction states. In both cases, the systems rely on learning-based methods to approximate these inverse mappings and on careful hardware design to ensure that the measured signals retain enough physical structure to support reliable inference.

6.2 Cross-System Comparison

A structured comparison across the three systems reveals both the diversity of design choices and the recurring trade-offs that characterize surface-integrated sensing. Table 6.1 summarizes the main technical dimensions for each system.

Table 6.1. *Cross-system comparison across the three surface-integrated sensing platforms.*

Dimension	Ch. 3: Vital Signs	Ch. 4: FBG E-skin	Ch. 5: Capacitive E-skin
Transduction	Optical (multi-λ PPG) + IR thermal	Optical (FBG wavelength shift)	Capacitive (field perturbation) + pneumatic pressure
Sensing channels	4 PPG λ + 768-pixel IR	21 FBG units	4 capacitive + 1 pressure
Sensing surface	Human body (fingertip, wrist)	Silicone-embedded fiber	TPU layer (16×16 cm) on 4 electrodes
Primary output	HR, SpO_2, Temperature, SBP/DBP	Touch location (x, y, z)	Object location (x, y, z), object count
Acquisition HW	MAX86916 EV + MLX90640 + Teensy 4.0	FBGs interrogator	FDC2214EVM + ABP2 + Teensy 4.1 (50 Hz)
ML approach	1D CNN architectures (5 families), ANN, PTT model	6-layer CNN regression	MLP surrogate + domain adaptation + permutation-invariant CNN/GRU
Deployment target	STM32N6 (INT8 quantized)	STM32F767ZI (compressed weights)	PC-side (offline)
Dataset source	Human subjects (88 for BP, 30 for temp.)	Robotic platform (2074 indentations)	UR5 robot (3 trajectory families, ~13,800 windows)
Key accuracy metric	SBP: ME 0.62±4.68 mmHg (INT8); Temp: MAE 0.11°C	MAE < 4 mm per axis; Euclidean ≈5 mm	Single-obj: MAE ≈10 mm (x,y), 2.6 mm (z); seq_gru: MAE ≈4.6 mm (x)

6.2.1 Transduction Principles and Signal Characteristics

The three systems exploit fundamentally different physical interactions, and these choices propagate through the entire pipeline. In Chapter 3, the transducer measures optical absorption modulated by blood-volume changes: the signal is physiologically generated, intrinsically tied to the cardiac cycle, and its quality depends on tissue properties and coupling pressure. In Chapter 4, the transducer measures wavelength shifts in Bragg gratings caused by strain propagation through a silicone matrix: the signal is mechanically generated and depends on material properties, indentation geometry, and the spatial relationship between contact and grating locations. In Chapter 5, the transducer measures electric field perturbation caused by the proximity or contact of an object: the signal depends on electrode-object geometry, dielectric properties, and the grounding environment.

A common feature is that none of these modalities provides a direct readout of the quantity of interest. PPG does not measure blood pressure; it provides waveform features from which BP must be inferred. FBG wavelength shifts do not directly encode contact coordinates; they provide distributed deformation patterns that must be decoded. Capacitive signals do not directly report object position; they provide field perturbation patterns that must be inverted. In all cases, the mapping is indirect and many-to-one, requiring either physics-based models (PTT for BP, Bragg equation for strain) or data-driven regression (CNN, MLP, GRU) to bridge the gap. The degree of ill-posedness, however, varies substantially: the capacitive system with only 5 channels faces the most severely under-constrained inverse problem, motivating the more elaborate learning pipeline developed in Chapter 5.

6.2.2 Electronics, Acquisition, and Hardware Stability

All three systems required non-trivial hardware engineering to produce signals that are stable enough for downstream learning. In Chapter 3, the acquisition chain relies on a commercial evaluation module (MAX86916) and a standardized optical stack. Variability is primarily biological (inter-subject differences in tissue, perfusion, pigmentation) rather than electronic, and the hardware challenge is to ensure consistent optical coupling through mechanical fixtures such as 3D-printed wrist support for the thermal camera. In Chapter 4, the FBG interrogator provides high-resolution wavelength readout, but the system-level challenge is embedding the fiber within a compliant silicone substrate while preserving strain transfer and avoiding mechanical degradation. In Chapter 5, the hardware challenge was the most acute: early prototypes suffered from parasitic capacitance, ground-shift artifacts, and unstable baselines caused by uncontrolled cable routing and grounding. The progressive refinement of shielding, cable management, and grounding strategy described in Section 5.2.3 was not a peripheral implementation detail but a prerequisite for producing physically meaningful signals.

This observation generalizes: in surface-integrated sensing, the dominant sources of measurement degradation are frequently at the system-integration level (parasitics, grounding, mechanical coupling) rather than at the transducer level. A sensor with excellent intrinsic sensitivity can still produce unusable data if its interconnections, shielding, or reference potentials are poorly managed. All three chapters implicitly confirm this point, but it is most visible in Chapter 5, where the electrical environment directly corrupted the spatial structure needed for learning.

6.2.3 Learning Pipelines and Data Strategies

The learning approaches adopted across the three systems reflect different data availability regimes and inverse-problem complexities. In Chapter 3, the edge-BP workflow compared five distinct 1D CNN architectures under a controlled benchmark protocol, with the primary research question being which architecture best survives quantization rather than which architecture achieves the lowest floating-point error. The dataset consisted of real PPG-BP pairs from human subjects, where ground truth is obtained from reference devices. The data strategy is conventional supervised regression with careful split design to avoid overlap leakage.

In Chapter 4, a single CNN architecture was trained to regress contact coordinates from FBG wavelength-shift windows. The dataset was collected under robotic control, which provided dense, repeatable ground-truth labels. The relatively high channel count (21 FBGs) and the localized nature of the sensing area made the inverse problem well-conditioned enough that a conventional supervised pipeline sufficed.

Chapter 5 required a substantially more complex learning framework, driven by two factors: the scarcity of annotated multi-object data and the severity of the inverse problem with only 5 channels. The solution was a simulation-assisted pipeline: a surrogate forward model learned from real single-object data, procedural synthesis of multi-object windows, domain adaptation to reduce the sim-to-real gap, and a permutation-invariant predictor. This escalation in methodological complexity is a direct consequence of the hardware constraints: fewer channels demand more sophisticated inference. Additionally, Chapter 5 introduced temporal modeling (GRU-based architectures) to exploit sequential structure in multi-object events, a dimension that was not necessary in the other two systems where interaction is either static (BP measurement window) or single-event (FBG indentation). These structural differences translate directly into localization performance. On real test data, the FBG system reached per-axis MAE values between 1.6 and 4.0 mm (Table 4.1), with a mean planar Euclidean error of about 5 mm, whereas the capacitive system, on its real single-object reference set, reached per-axis MAE values between 2.6 and 10.5 mm (Table 5.1). The FBG skin is thus consistently more accurate, but the gap reflects the conditioning of the inverse problem rather than the transduction principle: 21 channels over a small, contact-engaged substrate yield a well-posed problem, while five channels over a 16 × 16 cm surface, additionally required to resolve proximity and multiple objects, yield a far more under-constrained one. The comparison must also be read with care, since the substantially lower capacitive errors obtained in the sequential multi-object setting (down to about 4.6 mm on x, Table 5.5) are measured on synthetic data and on a different task and are therefore not directly comparable to the FBG real-data figures. The two skins are best understood as complementary points in the surface-sensing design space, dense, contact-only and well-conditioned versus sparse, proximity-and-contact and deliberately under-constrained, rather than as competing solutions to a single task.

A cross-cutting observation is that data quality and label reliability mattered at least as much as model architecture in all three systems. In Chapter 3, ground-truth heterogeneity across campaigns (sphygmomanometer vs pulse oximeter for HR) was explicitly flagged as a comparability issue. In Chapter 4, the robotic acquisition protocol ensured dense and reproducible ground truth, which is reflected in the relatively low localization error. In Chapter 5, the framework was designed to acknowledge label limitations: the force-related variable was retained only as an auxiliary conditioning quantity and was not evaluated as a primary output, precisely because the available annotation did not support an independent force estimation claim.

6.2.4 Embedded Deployment and Resource Constraints

The three systems illustrate different points along the deployment spectrum. Chapter 3 achieved full on-device inference on an STM32N6 microcontroller with INT8 quantization, demonstrating that a clinically relevant BP estimator can execute within hundreds of kilobytes of Flash and sub-100 KiB of RAM. The key insight was that quantization robustness is architecture-dependent: the compact residual design (Residual 1D CNN-Slim) preserved its error statistics under INT8 conversion, whereas the simpler baseline CNN degraded substantially[101]. This finding motivated the selection of the Slim architecture despite its larger absolute footprint.

Chapter 4 deployed its CNN on an STM32F767ZI using weight compression (4× reduction) and validated on target with a total inference time of approximately 12 ms. The numerical deviation introduced by deployment was minimal (RMSE 0.36 mm), confirming that the model architecture was already suitable for MCU-class execution. The system also demonstrated end-to-end real-time operation through UDP streaming and GUI feedback.

Chapter 5 did not include embedded deployment: the multi-object learning pipeline operates on a host workstation. This is a natural consequence of the pipeline complexity (surrogate model, domain adaptation, temporal encoding) and of the current focus on validating the sensing and learning methodology before addressing deployment. However, the single-object localization model and the baseline CNN are architecturally comparable to the models deployed in Chapters 3 and 4, suggesting that embedded deployment would be feasible for the simpler pipeline stages. Porting the full multi-object framework, including domain adaptation and GRU-based temporal encoding, to MCU-class hardware remains an open challenge that would require additional compression and architecture simplification.

6.3 Shared Methodological Lessons

Beyond the system-specific contributions, the three chapters converge on several methodological principles that generalize across surface-integrated sensing applications.

Hardware stability as a prerequisite for learning. In all three systems, the quality of the measured signal was not primarily limited by transducer sensitivity, but by integration-level factors: optical coupling and mechanical fixtures in Chapter 3, silicone-fiber bonding in Chapter 4, and grounding and parasitic management in Chapter 5. The implication is that, for surface-integrated sensing, hardware refinement is not an engineering afterthought to be handled before the "real" machine learning work begins; it is part of the scientific contribution, because the learning pipeline can only extract patterns that the hardware actually preserves. This is especially critical in sparse or indirect sensing, where corrupted signal structure cannot be compensated by more complex models.

Validation must match deployment conditions. A recurring theme is that offline accuracy on a workstation is not sufficient performance claim. Chapter 3 demonstrated that floating-point error is not a reliable predictor of INT8-quantized performance, and that deployment constraints (latency, memory, numerical stability) should be co-reported. Chapter 4 validated the CNN on the target MCU and measured inference time and numerical deviation. Chapter 5 adopted stratified evaluation by depth regime, revealing that aggregate metrics can hide systematic performance degradation in physically challenging conditions. Together, these practices suggest that validation in surface-integrated sensing should be multi-layered: accuracy under deployment constraints, error decomposition by operating regime, and explicit acknowledgment of ground-truth limitations.

Data-centric design over model-centric optimization. The three chapters placed greater emphasis on dataset construction, label quality, and split design than on exhaustive architectural search. In Chapter 3, careful separation of HR ground-truth sources and leakage removal after overlapping segmentation were essential to the validity of BP results. In Chapter 4, the robotic acquisition protocol provided dense, reproducible labels. In Chapter 5, the entire multi-object pipeline was motivated by the unavailability of annotated real multi-object data, leading to a simulation-assisted synthesis strategy. In all cases, the investment in data infrastructure produced more reliable outcomes than further model complexity would have. This observation is consistent with the growing recognition in the embedded ML community that data quality and curation are first-order concerns in resource-constrained settings[102], [103].

Co-design of sensing, representation, and inference. None of the three systems treated sensing hardware and machine learning as separable components. In Chapter 3, the choice of multi-wavelength PPG directly influenced feature extraction and artifact mitigation strategies. In Chapter 4, the spatial distribution of FBGs within the silicone matrix determined the deformation patterns available for learning. In Chapter 5, the electrode topology, grounding refinement, and baseline correction shaped the signal representation fed to the learning framework. The lesson is that, in surface-integrated sensing, the system boundary of the machine learning problem extends to the transducer, the substrate, and the electrical interface. Optimizing only the model while ignoring hardware constraints leads to performance claims that may not transfer to the real sensing chain.

6.4 Implications for Healthcare Robotics

The systems developed in this thesis were motivated by healthcare robotics scenarios, and the comparative discussion above provides a basis for assessing what the current results imply for this application domain.

A first observation is that surface-integrated sensing can serve multiple functions on the same robotic platform. A healthcare robot equipped with an electronic skin for contact and proximity detection (Chapters 4-5) could simultaneously host a vital-sign monitoring module (Chapter 3) on a dedicated interaction surface. These functions are complementary: the e-skin provides safety and interaction awareness during physical collaboration, while the physiological module supports health monitoring during structured contact. The co-existence of both capabilities on a single platform would create a more informative interaction loop, where the robot not only avoids harmful contact but also gathers clinically relevant data when appropriate contact occurs[104].

However, the current state of each system also reveals the distance from clinical deployment. The vital-sign monitoring module has been validated under controlled acquisition conditions with limited population diversity (primarily healthy subjects); generalization to pathological conditions, continuous operation, and fully on-device preprocessing remains to be demonstrated. The FBG-based e-skin has shown real-time localization on embedded hardware, but its sensing area is still limited, and the system has not been tested under the mechanical and hygienic constraints of a clinical environment. The capacitive-pneumatic e-skin has demonstrated the most complex inference capability (multi-object localization with temporal encoding), but its multi-object validation remains predominantly synthetic and its deployment is still offline.

For healthcare robotics, several additional requirements would need to be addressed. Hygiene compatibility demands surface materials and coatings that withstand frequent cleaning or disposable covers, without degrading sensing performance[105]. Long-term reliability requires characterization of drift, aging, and mechanical fatigue under realistic duty cycles. Safety certification involves not only collision detection but also demonstrating that sensing failures degrade gracefully rather than producing dangerous false negatives[106]. Privacy and data handling become relevant when physiological data are continuously acquired during interaction.

From a system-integration perspective, the heterogeneity of the three platforms (different microcontrollers, different communication protocols, different processing architectures) reflects the prototypal nature of the work. A unified platform for healthcare robotics would benefit from a common embedded architecture, shared data buses, and standardized interfaces that allow modular deployment and maintenance of different sensing skins. The edge-deployment results from Chapters 3 and 4 suggest that MCU-class hardware can support the computational requirements of individual sensing modalities; the challenge is to scale this to multi-modal operation while respecting the joint power, latency, and memory budget.

In summary, the present work demonstrates the feasibility of surface-integrated sensing across both physiological monitoring and physical interaction perception, and provides validated building blocks, including hardware, signal processing, learning frameworks, and deployment strategies, that are relevant to healthcare robotics. Bridging the gap between these prototypal demonstrations and clinical-grade deployment will require broader validation cohorts, long-term reliability studies, multi-modal integration on unified platforms, and explicit engagement with safety and regulatory requirements. These directions are developed further in the following chapter.

7 Conclusions and Future Work

Chapter 6 examined the three experimental systems developed in this thesis from a comparative and system-level perspective, distilling shared methodological lessons and discussing their collective implications for healthcare robotics. This closing chapter serves a different purpose. Rather than looking across the systems, it looks back at what each chapter concretely delivered, identifies limitations that constrain the scope of the claims supported by the thesis as a whole, and outlines research directions that follow from the contributions and limitations identified. The discussion is organized as follows: Section 7.1 summarizes the specific contributions of each chapter; Section 7.2 discusses limitations that apply to the thesis rather than to any individual chapter; and Section 7.3 outlines future directions.

7.1 Summary of Main Contributions

The work reported in this thesis delivered three experimental systems and a comparative framework, organized around the premise that surface-integrated sensing can support both physiological and interactional awareness in human-robot scenarios. The specific contributions of each chapter are summarized below, with emphasis on the concrete artifacts, datasets, models, and performance figures that distinguish them.

Chapter 3 developed a compact vital-sign monitoring pipeline combining multi-wavelength photoplethysmography and infrared thermometry. Its contributions were threefold: the definition of an acquisition and preprocessing chain adapted to finger-placed measurements; a set of estimation models for heart rate, oxygen saturation, skin temperature, and cuffless blood pressure; and the integration of the pipeline on the STM32N6 microcontroller, which demonstrated that multi-parameter physiological estimation can be carried out within the resource envelope of a single edge device.

Chapter 4 presented an electronic skin based on fiber Bragg grating sensors for indentation localization on a robot surface. Its contributions included the design and characterization of the FBG sensing layer; the acquisition of a 2074-sample indentation dataset using a bimanual robotic platform; a regression model that reached a mean absolute error below 4 mm and a mean Euclidean distance of approximately 5 mm; and the deployment of the model on the STM32F767ZI through X-CUBE-AI, which reached an inference time of about 12 ms with only 0.36 mm of RMSE degradation with respect to the floating-point reference.

Chapter 5 developed a capacitive-pneumatic electronic skin for proximity sensing and 3D object localization. Its contributions were multiple: the hardware refinement that enabled stable acquisition on a four-electrode plus pressure architecture, including grounding, shielding, and the elimination of the ground-shift phenomenon; a robotic acquisition campaign covering grid, parabolic, and random trajectories; a learning framework that trained a surrogate MLP on real single-object data, used it to synthesize multi-object interactions, and applied domain adaptation together with a permutation-invariant loss to predict up to two concurrent objects; a temporal extension based on GRUs that reduced the mean absolute error from approximately 13.7 mm to approximately 4.6 mm on sequential multi-object events; and the identification of mixed-regime training as the most robust compromise across single- and multi-object conditions.

Finally, **Chapter 6** provided a contribution of a different nature. Rather than introducing a new system, it reframed the three experimental platforms as instances of a single design space and extracted methodological observations that would not have been visible from any individual chapter taken in isolation. This cross-system reading is itself part of the thesis contribution, and the fact that it could be written at a consistent level of abstraction across otherwise very different sensing modalities is one of the main arguments supporting the unifying framing adopted from the outset.

Read against the introduction, these contributions also speak to the four cross-cutting challenges identified at the outset: low-power electronics, robustness to motion and variability, signal quality assessment, and learning models that generalize across users and conditions. The first was directly addressed, since two of the three systems were brought to microcontroller-class deployment (the vital-sign pipeline on the STM32N6 with INT8 quantization, and the FBG skin on the STM32F767ZI), showing that the proposed methods fit within tight power and memory envelopes. Signal quality assessment was likewise treated as a first-class concern rather than a preprocessing afterthought, most visibly in the grounding and shielding refinement that conditioned the capacitive measurements (Section 5.2.3) and in the use of quantization robustness as a model-selection criterion (Chapter 3). The remaining two challenges were only partially met: robustness to motion and variability was examined under controlled laboratory conditions but not under the unconstrained motion and inter-subject variability of real deployment, and generalization across users and conditions was constrained by the limited cohorts and single experimental platforms used. These last two challenges therefore frame the limitations and future directions discussed in the following sections, connecting the outcomes of the thesis back to the agenda set in the introduction.

7.2 Limitations

The contributions above were obtained within a constrained experimental scope. The limitations listed here are deliberately distinct from the chapter-specific caveats discussed at the end of Chapters 3, 4, and 5, and from the deployment-oriented considerations raised in Section 6.4: they concern the boundary of what the thesis as a whole can and cannot establish, rather than the engineering constraints of any individual system or the clinical readiness of the overall direction.

The first limitation concerns the statistical diversity of experimental evidence. The vital-sign pipeline was evaluated on a limited cohort, the FBG e-skin was tested on a single bimanual platform, and the capacitive e-skin was tested on a single UR5 cell. Although the pipelines were designed to be transferable, the breadth of conditions required to support claims of population-level generalization, across age groups, or skin tones, was not covered[107], [108]. The figures reported in this thesis should therefore be read as evidence that each approach is feasible and that its internal mechanisms behave as intended, not as an estimate of the accuracy expected in an arbitrary target population.

A second limitation concerns the scope of methodological exploration. For each learning problem, the thesis committed to a specific family of models and to a specific compression and deployment path, CNN for the FBG indentation regression, a surrogate MLP combined with a domain-adaptation stage and a permutation-invariant head for the capacitive e-skin, and lightweight regressors for the physiological pipeline, all compressed and quantized for microcontroller deployment. Alternative paradigms, such as transformer-based temporal models[109], physics-informed networks[110], or self-supervised pretraining on unlabeled sensor streams[111], were not systematically evaluated. This was a deliberate choice driven by the embedded-target constraints, but it means that the performance figures reported here are informative about the chosen paradigms rather than about the upper bound achievable on each task.

A third limitation concerns the scale of the datasets used. The FBG indentation dataset contains 2074 samples, and the real portion of the capacitive e-skin dataset is single-object only, with multi-object supervision entirely synthetic. These scales were sufficient to support the architectural and methodological arguments made in the respective chapters, and the synthesis pipeline of Chapter 5 was explicitly designed to compensate for the absence of real multi-object data, but they do not support strong claims about long-tail robustness, rare-event behavior, or the extrapolation properties of the models outside their training distribution.

A fourth limitation is temporal. The thesis provides a snapshot rather than a longitudinal evaluation. Sensor drift, long-term calibration stability, re-training requirements after hardware modifications, and wear-related degradation of the electronic skins were not quantified. For systems intended to operate continuously, and in particular in healthcare environments, these aspects are ultimately as relevant as the point-in-time accuracy figures reported here, and their characterization is part of the work that must precede any move beyond controlled laboratory validation.

7.3 Future Directions

Several concrete research directions follow from the contributions and limitations identified above. They are listed in approximate order of maturity, beginning with extensions that can be pursued with the existing hardware and proceeding toward longer-horizon research programs. These directions are intentionally specific and actionable and should be read as complementary to the higher-level deployment considerations discussed in Section 6.4.

The most immediate priority is to close the synthetic-to-real gap for multi-object capacitive sensing. The procedure established in Chapter 5 produced strong synthetic-domain results but cannot, on its own, establish real-world multi-object performance. A targeted acquisition campaign using two physical objects on the UR5 cell, following the trajectory taxonomy already developed, would allow the permutation-invariant and GRU-based models to be evaluated against ground-truth labels rather than generated ones, and would enable a direct quantification of the sim-to-real gap that the current results can only bound indirectly[112].

For the vital-sign pipeline, the natural next step is a structured clinical validation. This would involve expanding the acquisition cohort along dimensions that are known to affect photoplethysmography and infrared thermometry (age, skin pigmentation[107], [108], [113], vasoactive medication) evaluating robustness to acquisition conditions that mimic realistic clinical workflows, including brief hand movements and partial contact. The STM32N6-based deployment developed in Chapter 3 provides an appropriate platform for such a campaign, as it already enables repeated and standardized acquisition.

A second family of directions concerns enriching the sensing capabilities of each system while preserving its embedded constraints. For the FBG e-skin, promising extensions include simultaneous indentation-and-force estimation and the use of temporal models to capture dynamic contact transitions rather than static indentations. For the capacitive e-skin, incorporating material or texture discrimination, possibly through controlled modulation of the pneumatic channel, would broaden the interaction repertoire beyond spatial localization. For the vital-sign pipeline, the integration of additional physiological channels such as electrodermal activity or respiration[114] could support a richer user-state model that goes beyond cardiovascular and thermal parameters.

A third direction concerns learning behavior after deployment. The current pipelines are trained offline and frozen before being flashed to the target microcontroller; they do not exploit the information that continues to arrive during operational use. On-device adaptation, ranging from simple recalibration of the output head to more structured continual learning schemes[115], [116], is a promising but largely unexplored avenue, particularly for the capacitive and FBG skins, where environmental changes such as humidity, temperature, and mechanical aging can induce measurable drift. The central challenge is to design adaptation mechanisms that preserve the safety and calibration properties validated at training time, an aspect that becomes particularly delicate when the same device is intended for use across multiple subjects or deployment cycles.

A fourth direction is the systematic evaluation of alternative learning paradigms on the tasks defined here. Transformer-based temporal models[109], self-supervised pretraining on large unlabeled sensor corpora[111], and physics-informed regularization[110] are all candidates for yielding stronger accuracy or better data efficiency, particularly for the capacitive e-skin where real multi-object supervision is scarce. The embedded-deployment constraint would have to be re-examined in light of these alternatives, but the emerging literature on efficient attention and on distilled small models suggests that the gap between strong offline models and deployable on-device models is narrowing rapidly[117], and that the trade-offs identified in this thesis may look different within a few model generations.

Finally, a longer-horizon direction is the user-centered study of these systems in realistic assistive and caregiving scenarios. The technical feasibility established in this thesis is a necessary but not sufficient condition for adoption; understanding how the systems are used by clinicians, caregivers, and patients, and how their outputs are interpreted and acted upon[118], [119], is the step that will ultimately determine whether surface-integrated sensing matures from a family of research prototypes into a practical component of the next generation of healthcare robotic systems.

Bibliography

[1] W. Heng, S. Solomon, and W. Gao, "Flexible Electronics and Devices as Human–Machine Interfaces for Medical Robotics," *Adv. Mater.*, vol. 34, no. 16, p. 2107902, Apr. 2022, doi: 10.1002/adma.202107902.

[2] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile Sensing—From Humans to Humanoids," *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 1–20, Feb. 2010, doi: 10.1109/TRO.2009.2033627.

[3] F. Liu, S. Deswal, A. Christou, Y. Sandamirskaya, M. Kaboli, and R. Dahiya, "Neuro-inspired electronic skin for robots," *Sci. Robot.*, vol. 7, no. 67, p. eabl7344, Jun. 2022, doi: 10.1126/scirobotics.abl7344.

[4] C. Xu, S. A. Solomon, and W. Gao, "Artificial intelligence-powered electronic skin," *Nat. Mach. Intell.*, vol. 5, no. 12, pp. 1344–1355, Dec. 2023, doi: 10.1038/s42256-023-00760-z.

[5] H. Oh, C. Lee, N. Kim, T. An, and G. Kim, "Review: Sensors for Biosignal/Health Monitoring in Electronic Skin," *Polymers*, vol. 13, no. 15, p. 2478, Jul. 2021, doi: 10.3390/polym13152478.

[6] X. Wang *et al.*, "Multistage Microstructured Ionic Skin for Real-Time Vital Signs Monitoring and Human-Machine Interaction," *ENERGY Environ. Mater.*, p. e12767, May 2024, doi: 10.1002/eem2.12767.

[7] Y. Zhou, J. Zhao, P. Lu, Z. Wang, and B. He, "TacSuit: A Wearable Large-Area, Bioinspired Multimodal Tactile Skin for Collaborative Robots," *IEEE Trans. Ind. Electron.*, vol. 71, no. 2, pp. 1708–1717, Feb. 2024, doi: 10.1109/TIE.2023.3253921.

[8] J. Tu *et al.*, "Electronic skins with multimodal sensing and perception," *Soft Sci.*,

vol. 3, no. 3, Jul. 2023, doi: 10.20517/ss.2023.15.
[9] A.-C. Bunea *et al.*, "E-Skin: The Dawn of a New Era of On-Body Monitoring Systems," *Micromachines*, vol. 12, no. 9, p. 1091, Sep. 2021, doi: 10.3390/mi12091091.
[10] J. R. Sempionatto *et al.*, "An epidermal patch for the simultaneous monitoring of haemodynamic and metabolic biomarkers," *Nat. Biomed. Eng.*, vol. 5, no. 7, pp. 737–748, Feb. 2021, doi: 10.1038/s41551-021-00685-1.
[11] F. Khoshmanesh, P. Thurgood, E. Pirogova, S. Nahavandi, and S. Baratchi, "Wearable sensors: At the frontier of personalised health monitoring, smart prosthetics and assistive technologies," *Biosens. Bioelectron.*, vol. 176, p. 112946, Mar. 2021, doi: 10.1016/j.bios.2020.112946.
[12] M. Hussain, M. O'Nils, J. Lundgren, and S. J. Mousavirad, "A Comprehensive Review on Deep Learning-Based Data Fusion," *IEEE Access*, vol. 12, pp. 180093–180124, 2024, doi: 10.1109/ACCESS.2024.3508271.
[13] E. Covi *et al.*, "Adaptive Extreme Edge Computing for Wearable Devices," *Front. Neurosci.*, vol. 15, p. 611300, May 2021, doi: 10.3389/fnins.2021.611300.
[14] T. Arnold and M. Scheutz, "The Tactile Ethics of Soft Robotics: Designing Wisely for Human–Robot Interaction," *Soft Robot.*, vol. 4, no. 2, pp. 81–87, Jun. 2017, doi: 10.1089/soro.2017.0032.
[15] A. Cherubini and D. Navarro-Alarcon, "Sensor-Based Control for Collaborative Robots: Fundamentals, Challenges, and Opportunities," *Front. Neurorobotics*, vol. 14, p. 576846, Jan. 2021, doi: 10.3389/fnbot.2020.576846.
[16] G. Pang, G. Yang, and Z. Pang, "Review of Robot Skin: A Potential Enabler for Safe Collaboration, Immersive Teleoperation, and Affective Interaction of Future Collaborative Robots," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 3, pp. 681–700, Aug. 2021, doi: 10.1109/TMRB.2021.3097252.
[17] F. Vicentini, "Collaborative Robotics: A Survey," *J. Mech. Des.*, vol. 143, no. 4, p. 040802, Apr. 2021, doi: 10.1115/1.4046238.
[18] A. Slepyan and N. Thakor, "Towards scalable soft e-skin: Flexible event-based tactile-sensors using wireless sensor elements embedded in soft elastomer," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, New York City, NY, USA: IEEE, Nov. 2020, pp. 334–339. doi: 10.1109/BioRob49111.2020.9224353.
[19] J. A. Barreiros *et al.*, "Haptic perception using optoelectronic robotic flesh for embodied artificially intelligent agents," *Sci. Robot.*, vol. 7, no. 67, p. eabi6745, Jun. 2022, doi: 10.1126/scirobotics.abi6745.
[20] C. C. Vu, S. J. Kim, and J. Kim, "Flexible wearable sensors - an update in view of touch-sensing," *Sci. Technol. Adv. Mater.*, vol. 22, no. 1, pp. 26–36, Dec. 2021, doi: 10.1080/14686996.2020.1862629.
[21] X. Lin *et al.*, "Biocompatible Multifunctional E-Skins with Excellent Self-Healing Ability Enabled by Clean and Scalable Fabrication," *Nano-Micro Lett.*, vol. 13, no. 1, p. 200, Dec. 2021, doi: 10.1007/s40820-021-00701-8.
[22] X. Peng *et al.*, "A breathable, biodegradable, antibacterial, and self-powered electronic skin based on all-nanofiber triboelectric nanogenerators," *Sci. Adv.*, vol. 6, no. 26, p. eaba9624, Jun. 2020, doi: 10.1126/sciadv.aba9624.
[23] F. Li *et al.*, "A polyurethane integrating self-healing, anti-aging and controlled degradation for durable and eco-friendly E-skin," *Chem. Eng. J.*, vol. 410, p. 128363, Apr. 2021, doi: 10.1016/j.cej.2020.128363.
[24] Z. Bai *et al.*, "Mechanically Robust and Transparent Organohydrogel-Based E-

Skin Nanoengineered from Natural Skin," *Adv. Funct. Mater.*, vol. 33, no. 15, p. 2212856, Apr. 2023, doi: 10.1002/adfm.202212856.

[25] M. Sang, K. Kim, J. Shin, and K. J. Yu, "Ultra-Thin Flexible Encapsulating Materials for Soft Bio-Integrated Electronics," *Adv. Sci.*, vol. 9, no. 30, p. 2202980, Oct. 2022, doi: 10.1002/advs.202202980.

[26] A. Dos Santos, E. Fortunato, R. Martins, H. Águas, and R. Igreja, "Transduction Mechanisms, Micro-Structuring Techniques, and Applications of Electronic Skin Pressure Sensors: A Review of Recent Advances," *Sensors*, vol. 20, no. 16, p. 4407, Aug. 2020, doi: 10.3390/s20164407.

[27] H. Xie, A. Jiang, L. Seneviratne, and K. Althoefer, "Pixel-based optical fiber tactile force sensor for robot manipulation," in *2012 IEEE Sensors*, Taipei, Taiwan: IEEE, Oct. 2012, pp. 1–4. doi: 10.1109/ICSENS.2012.6411462.

[28] M. G. S. Murshed, C. Murphy, D. Hou, N. Khan, G. Ananthanarayanan, and F. Hussain, "Machine Learning at the Network Edge: A Survey," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–37, Nov. 2022, doi: 10.1145/3469029.

[29] Y. Wang, Y. Yang, and Z. L. Wang, "Triboelectric nanogenerators as flexible power sources," *Npj Flex. Electron.*, vol. 1, no. 1, p. 10, Nov. 2017, doi: 10.1038/s41528-017-0007-8.

[30] J. Wang, M. D. Alea, J. Van Assche, and G. G. E. Gielen, "End-to-End Optimization of High-Density e-Skin Design: From Spiking Taxel Readout to Texture Classification," in *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Antwerp, Belgium: IEEE, Apr. 2023, pp. 1–6. doi: 10.23919/DATE56975.2023.10137077.

[31] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden: IEEE, May 2016, pp. 536–543. doi: 10.1109/ICRA.2016.7487176.

[32] F. Sakr, F. Bellotti, R. Berta, and A. De Gloria, "Machine Learning on Mainstream Microcontrollers," *Sensors*, vol. 20, no. 9, p. 2638, May 2020, doi: 10.3390/s20092638.

[33] S. A. R. Zaidi, A. M. Hayajneh, M. Hafeez, and Q. Z. Ahmed, "Unlocking Edge Intelligence Through Tiny Machine Learning (TinyML)," *IEEE Access*, vol. 10, pp. 100867–100877, 2022, doi: 10.1109/ACCESS.2022.3207200.

[34] F. Hohman, M. B. Kery, D. Ren, and D. Moritz, "Model Compression in Practice: Lessons Learned from Practitioners Creating On-device Machine Learning Experiences," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–18. doi: 10.1145/3613904.3642109.

[35] S. S. Saha, S. S. Sandha, and M. Srivastava, "Machine Learning for Microcontroller-Class Hardware: A Review," *IEEE Sens. J.*, vol. 22, no. 22, pp. 21362–21390, Nov. 2022, doi: 10.1109/JSEN.2022.3210773.

[36] F. Cini, T. Banfi, G. Ciuti, L. Craighero, and M. Controzzi, "The relevance of signal timing in human-robot collaborative manipulation," *Sci. Robot.*, vol. 6, no. 58, p. eabg1308, Sep. 2021, doi: 10.1126/scirobotics.abg1308.

[37] G. Li, S. Liu, L. Wang, and R. Zhu, "Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition," *Sci. Robot.*, vol. 5, no. 49, p. eabc8134, Dec. 2020, doi: 10.1126/scirobotics.abc8134.

[38] K. Park, H. Yuk, M. Yang, J. Cho, H. Lee, and J. Kim, "A biomimetic elastomeric robot skin using electrical impedance and acoustic tomography for tactile sensing,"

Sci. Robot., vol. 7, no. 67, p. eabm7187, Jun. 2022, doi: 10.1126/scirobotics.abm7187.

[39] C. Botrugno, E. Leogrande, F. Dell'Olio, and T. Natale, "Smart combination of ECG and PPG signals: an innovative approach towards an electronic device for vital signs monitoring," in *2023 9th International Workshop on Advances in Sensors and Interfaces (IWASI)*, Monopoli (Bari), Italy: IEEE, Jun. 2023, pp. 256–260. doi: 10.1109/IWASI58316.2023.10164373.

[40] E. Leogrande, C. Botrugno, T. Natale, and F. Dell'Olio, "Contact-Less Body Temperature Monitoring by Infrared Camera: Accuracy Preliminary Assessment," in *2023 IEEE SENSORS*, Vienna, Austria: IEEE, Oct. 2023, pp. 1–4. doi: 10.1109/SENSORS56945.2023.10325125.

[41] C. Botrugno, E. Leogrande, T. Natale, and F. Dell'Olio, "A Multi-parameter Sensing Device for Vital Signs Monitoring," in *Proceedings of SIE 2024*, vol. 1263, M. Valle, P. Gastaldo, and E. Limiti, Eds., in Lecture Notes in Electrical Engineering, vol. 1263. , Cham: Springer Nature Switzerland, 2025, pp. 429–434. doi: 10.1007/978-3-031-71518-1_51.

[42] E. Leogrande *et al.*, "A multiwavelength photoplethysmography dataset with blood pressure and heart rate reference measurements," *Data Brief*, vol. 63, p. 112284, Dec. 2025, doi: 10.1016/j.dib.2025.112284.

[43] E. Leogrande, E. D. Luca, and F. Dell'Olio, "From PPG to Blood Pressure at the Edge: Quantization-Aware Architecture Selection and On-MCU Validation," *Sensors*, vol. 26, no. 9, p. 2674, Apr. 2026, doi: 10.3390/s26092674.

[44] S. C. Mathews, M. J. McShea, C. L. Hanley, A. Ravitz, A. B. Labrique, and A. B. Cohen, "Digital health: a path to validation," *Npj Digit. Med.*, vol. 2, no. 1, p. 38, May 2019, doi: 10.1038/s41746-019-0111-3.

[45] F. Fatehi, M. Samadbeik, and A. Kazemi, "What is Digital Health? Review of Definitions," in *Studies in Health Technology and Informatics*, A. Värri, J. Delgado, P. Gallos, M. Hägglund, K. Häyrinen, U.-M. Kinnunen, L. B. Pape-Haugaard, L.-M. Peltonen, K. Saranto, and P. Scott, Eds., IOS Press, 2020. doi: 10.3233/SHTI200696.

[46] D. Ray, T. Collins, S. Woolley, and P. Ponnapalli, "A Review of Wearable Multi-Wavelength Photoplethysmography," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 136–151, 2023, doi: 10.1109/RBME.2021.3121476.

[47] T. Tamura, "Current progress of photoplethysmography and SPO2 for health monitoring," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 21–36, Feb. 2019, doi: 10.1007/s13534-019-00097-w.

[48] S. Alharbi *et al.*, "Oxygen Saturation Measurements from Green and Orange Illuminations of Multi-Wavelength Optoelectronic Patch Sensors," *Sensors*, vol. 19, no. 1, p. 118, Dec. 2018, doi: 10.3390/s19010118.

[49] N. F. A. Jamal and K. A. Sidek, "Portable health monitoring kit using photolethysmogram signal," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 15, no. 2, p. 638, Aug. 2019, doi: 10.11591/ijeecs.v15.i2.pp638-649.

[50] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.

[51] K. Natarajan *et al.*, "Photoplethysmography Fast Upstroke Time Intervals Can Be Useful Features for Cuff-Less Measurement of Blood Pressure Changes in Humans," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 1, pp. 53–62, Jan. 2022, doi: 10.1109/TBME.2021.3087105.

[52] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007, doi: 10.1088/0967-3334/28/3/R01.

[53] V. S. Roha, R. Ranjan, and M. R. Yuce, "Evolving Blood Pressure Estimation: From Feature Analysis to Image-Based Deep Learning Models," *J. Med. Syst.*, vol. 49, no. 1, p. 97, Jul. 2025, doi: 10.1007/s10916-025-02228-6.

[54] Kejia Li and S. Warren, "Principle component analysis on photoplethysmograms: Blood oxygen saturation estimation and signal segmentation," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA: IEEE, Aug. 2011, pp. 7171–7174. doi: 10.1109/IEMBS.2011.6091812.

[55] Parry Fung, G. Dumont, C. Ries, C. Mott, and M. Ansermino, "Continuous noninvasive blood pressure measurement by pulse transit time," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, CA, USA: IEEE, 2004, pp. 738–741. doi: 10.1109/IEMBS.2004.1403264.

[56] D. Bernard, C. Msigwa, and J. Yun, "Toward IoT-Based Medical Edge Devices: PPG-Based Blood Pressure Estimation Application," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 5240–5255, Mar. 2023, doi: 10.1109/JIOT.2022.3222477.

[57] E. Leogrande, F. Dell'Olio, S. Mazzoleni, C. M. Oddo, and M. Filosa, "Edge AI Algorithm for FBG-Based E-Skin Touch Localization on Embedded Electronics," in *2024 IEEE SENSORS*, Kobe, Japan: IEEE, Oct. 2024, pp. 1–4. doi: 10.1109/SENSORS60989.2024.10784928.

[58] E. Leogrande *et al.*, "Electronic skin technologies: From hardware building blocks and tactile sensing to control algorithms and applications," *Sens. Actuators Rep.*, vol. 9, p. 100312, Jun. 2025, doi: 10.1016/j.snr.2025.100312.

[59] F. Sherwani, M. M. Asad, and B. S. K. K. Ibrahim, "Collaborative Robots and Industrial Revolution 4.0 (IR 4.0)," in *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, Karachi, Pakistan: IEEE, Mar. 2020, pp. 1–5. doi: 10.1109/ICETST49965.2020.9080724.

[60] A. Pauliková, Z. Gyurák Babeľová, and M. Ubárová, "Analysis of the Impact of Human–Cobot Collaborative Manufacturing Implementation on the Occupational Health and Safety and the Quality Requirements," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 4, p. 1927, Feb. 2021, doi: 10.3390/ijerph18041927.

[61] R. Galin and R. Meshcheryakov, "Automation and robotics in the context of Industry 4.0: the shift to collaborative robots," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 537, no. 3, p. 032073, May 2019, doi: 10.1088/1757-899X/537/3/032073.

[62] L. Massari *et al.*, "A Machine-Learning-Based Approach to Solve Both Contact Location and Force in Soft Material Tactile Sensors," *Soft Robot.*, vol. 7, no. 4, pp. 409–420, Aug. 2020, doi: 10.1089/soro.2018.0172.

[63] L. Massari *et al.*, "Functional mimicry of Ruffini receptors with fibre Bragg gratings and deep neural networks enables a bio-inspired large-area tactile-sensitive skin," *Nat. Mach. Intell.*, vol. 4, no. 5, pp. 425–435, May 2022, doi: 10.1038/s42256-022-00487-3.

[64] S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, "The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 12861–12885, Aug. 2022, doi: 10.1109/JIOT.2021.3139827.

[65] J. Tegin and J. Wikander, "Tactile sensing in intelligent robotic manipulation – a

review," *Ind. Robot Int. J.*, vol. 32, no. 1, pp. 64–70, Feb. 2005, doi: 10.1108/01439910510573318.

[66] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Substantial capabilities of robotics in enhancing industry 4.0 implementation," *Cogn. Robot.*, vol. 1, pp. 58–75, 2021, doi: 10.1016/j.cogr.2021.06.001.

[67] R. Islam *et al.*, "The Future of Cloud Computing: Benefits and Challenges," *Int. J. Commun. Netw. Syst. Sci.*, vol. 16, no. 04, pp. 53–65, 2023, doi: 10.4236/ijcns.2023.164004.

[68] G. Plastiras, M. Terzi, C. Kyrkou, and T. Theocharides, "Edge Intelligence: Challenges and Opportunities of Near-Sensor Machine Learning Applications," in *2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, Milan, Italy: IEEE, Jul. 2018, pp. 1–7. doi: 10.1109/ASAP.2018.8445118.

[69] Z. Zhang and J. Li, "A Review of Artificial Intelligence in Embedded Systems," *Micromachines*, vol. 14, no. 5, p. 897, Apr. 2023, doi: 10.3390/mi14050897.

[70] Dr. L. Dutta and S. Bharali, "TinyML Meets IoT: A Comprehensive Survey," *Internet Things*, vol. 16, p. 100461, Dec. 2021, doi: 10.1016/j.iot.2021.100461.

[71] R. Suresh, S. C. Tjin, and J. Hao, "Fiber Bragg Grating," in *Smart Materials in Structural Health Monitoring, Control and Biomechanics*, in Advanced Topics in Science and Technology in China. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 413–439. doi: 10.1007/978-3-642-24463-6_11.

[72] B. Da Silva Falcão *et al.*, "Strain measurement with multiplexed FBG sensor arrays: An experimental investigation," *Heliyon*, vol. 9, no. 8, p. e18652, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18652.

[73] R. Immonen and T. Hämäläinen, "Tiny Machine Learning for Resource-Constrained Microcontrollers," *J. Sens.*, vol. 2022, pp. 1–11, Nov. 2022, doi: 10.1155/2022/7437023.

[74] W. Xiang, K. Yu, F. Han, L. Fang, D. He, and Q.-L. Han, "Advanced Manufacturing in Industry 5.0: A Survey of Key Enabling Technologies and Future Trends," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 1055–1068, Feb. 2024, doi: 10.1109/TII.2023.3274224.

[75] R. Schiavi, A. Bicchi, and F. Flacco, "Integration of active and passive compliance control for safe human-robot coexistence," in *2009 IEEE International Conference on Robotics and Automation*, Kobe: IEEE, May 2009, pp. 259–264. doi: 10.1109/ROBOT.2009.5152571.

[76] T. S. Tadele, T. De Vries, and S. Stramigioli, "The Safety of Domestic Robotics: A Survey of Various Safety-Related Publications," *IEEE Robot. Autom. Mag.*, vol. 21, no. 3, pp. 134–142, Sep. 2014, doi: 10.1109/MRA.2014.2310151.

[77] S. E. Navarro *et al.*, "Proximity Perception in Human-Centered Robotics: A Survey on Sensing Systems and Applications," *IEEE Trans. Robot.*, vol. 38, no. 3, pp. 1599–1620, Jun. 2022, doi: 10.1109/TRO.2021.3111786.

[78] D. Goger, H. Alagi, and H. Worn, "Tactile proximity sensors for robotic applications," in *2013 IEEE International Conference on Industrial Technology (ICIT)*, Cape Town: IEEE, Feb. 2013, pp. 978–983. doi: 10.1109/ICIT.2013.6505804.

[79] H. Alagi, S. E. Navarro, M. Mende, and B. Hein, "A versatile and modular capacitive tactile proximity sensor," in *2016 IEEE Haptics Symposium (HAPTICS)*, Philadelphia, PA: IEEE, Apr. 2016, pp. 290–296. doi: 10.1109/HAPTICS.2016.7463192.

[80] Z. Liu, D. Chen, J. Ma, D. Jia, T. Wang, and Q. Su, "A Self-Capacitance Proximity E-Skin With Long-Range Sensibility for Robotic Arm Environment Perception," *IEEE Sens. J.*, vol. 23, no. 13, pp. 14854–14863, Jul. 2023, doi: 10.1109/JSEN.2023.3275206.

[81] C. Tang, X. Chang, J. Wang, Y. Peng, H. Wu, and H. Wang, "A Nonarray Soft Capacitive Tactile Sensor With Simultaneous Contact Force and Location Measurement for Intelligent Robotic Grippers," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–9, 2024, doi: 10.1109/TIM.2023.3343795.

[82] X. Xu, X. Guo, W. Zheng, and H. Liu, "A Flexible Electrical Capacitance Tomography Sensor for Proximity Object Recognition," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025, doi: 10.1109/TIM.2025.3551985.

[83] N. Shrivastava, R. Mudumbai, U. Madhow, and S. Suri, "Target tracking with binary proximity sensors," *ACM Trans. Sens. Netw.*, vol. 5, no. 4, pp. 1–33, Nov. 2009, doi: 10.1145/1614379.1614382.

[84] H. Sun and G. Martius, "Machine Learning for Haptics: Inferring Multi-Contact Stimulation From Sparse Sensor Configuration," *Front. Neurorobotics*, vol. 13, p. 51, 2019, doi: 10.3389/fnbot.2019.00051.

[85] H. Sun, A. Spiers, H. Lee, J. Fiene, and G. Martius, "Sensing multi-directional forces at superresolution using taxel value isoline theory," *Nat. Commun.*, vol. 16, no. 1, p. 8031, Aug. 2025, doi: 10.1038/s41467-025-63230-7.

[86] D. Kong *et al.*, "Super-resolution tactile sensor arrays with sparse units enabled by deep learning," *Sci. Adv.*, vol. 11, no. 27, p. eadv2124, Jul. 2025, doi: 10.1126/sciadv.adv2124.

[87] A. B. Dawood, C. Coppola, and K. Althoefer, "Learning Decoupled Multi-touch Force Estimation, Localization and Stretch for Soft Capacitive E-skin," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, United Kingdom: IEEE, May 2023, pp. 614–619. doi: 10.1109/ICRA48891.2023.10160961.

[88] Z. Jiang, Z. Xu, M. Li, H. Zeng, F. Gong, and Y. Tang, "Electrical Impedance Tomography-Based Electronic Skin for Multi-Touch Tactile Sensing Using Hydrogel Material and FISTA Algorithm," *Sensors*, vol. 24, no. 18, p. 5985, Sep. 2024, doi: 10.3390/s24185985.

[89] H. Park, K. Park, S. Mo, and J. Kim, "Deep Neural Network Based Electrical Impedance Tomographic Sensing Methodology for Large-Area Robotic Tactile Sensing," *IEEE Trans. Robot.*, vol. 37, no. 5, pp. 1570–1583, Oct. 2021, doi: 10.1109/TRO.2021.3060342.

[90] Z. Husain, N. A. Madjid, and P. Liatsis, "Tactile Sensing Using Machine Learning-Driven Electrical Impedance Tomography," *IEEE Sens. J.*, vol. 21, no. 10, pp. 11628–11642, May 2021, doi: 10.1109/JSEN.2021.3054870.

[91] R. Patel, R. Cox, and N. Correll, "Integrated proximity, contact and force sensing using elastomer-embedded commodity proximity sensors," *Auton. Robots*, vol. 42, no. 7, pp. 1443–1458, Oct. 2018, doi: 10.1007/s10514-018-9751-4.

[92] S. Taylor, K. Park, S. Yamsani, and J. Kim, "Fully 3D printable Robot Hand and Soft Tactile Sensor based on Air-pressure and Capacitive Proximity Sensing," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan: IEEE, May 2024, pp. 18100–18105. doi: 10.1109/ICRA57147.2024.10610731.

[93] S. Zhong, A. Albini, P. Maiolino, and I. Posner, "TactGen: Tactile Sensory Data Generation via Zero-Shot Sim-to-Real Transfer," *IEEE Trans. Robot.*, vol. 41, pp.

1316–1328, 2025, doi: 10.1109/TRO.2024.3521967.
[94] H. Lee *et al.*, “Predicting the Force Map of an ERT-Based Tactile Sensor Using Simulation and Deep Networks,” *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 1, pp. 425–439, Jan. 2023, doi: 10.1109/TASE.2022.3156184.
[95] Y. Narang, B. Sundaralingam, M. Macklin, A. Mousavian, and D. Fox, “Sim-to-Real for Robotic Tactile Sensing via Physics-Based Simulation and Learned Latent Projections,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi’an, China: IEEE, May 2021, pp. 6444–6451. doi: 10.1109/ICRA48506.2021.9561969.
[96] H. Lee, H. Park, G. Serhat, H. Sun, and K. J. Kuchenbecker, “Calibrating a Soft ERT-Based Tactile Sensor with a Multiphysics Model and Sim-to-real Transfer Learning,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France: IEEE, May 2020, pp. 1632–1638. doi: 10.1109/ICRA40945.2020.9196732.
[97] D. F. Gomes, P. Paoletti, and S. Luo, “Generation of GelSight Tactile Images for Sim2Real Learning,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 4177–4184, Apr. 2021, doi: 10.1109/LRA.2021.3063925.
[98] Z. Si and W. Yuan, “Taxim: An Example-Based Simulation Model for GelSight Tactile Sensors,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2361–2368, Apr. 2022, doi: 10.1109/LRA.2022.3142412.
[99] W. Chen *et al.*, “Bidirectional Sim-to-Real Transfer for GelSight Tactile Sensors With CycleGAN,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6187–6194, Jul. 2022, doi: 10.1109/LRA.2022.3167064.
[100] R. Boldt, J. Gogulski, J. Gúzman-Lopéz, S. Carlson, and A. Pertovaara, “Two-point tactile discrimination ability is influenced by temporal features of stimulation,” *Exp. Brain Res.*, vol. 232, no. 7, pp. 2179–2185, Jul. 2014, doi: 10.1007/s00221-014-3908-y.
[101] P. P. Ray, “A review on TinyML: State-of-the-art and prospects,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1595–1623, Apr. 2022, doi: 10.1016/j.jksuci.2021.11.019.
[102] D. Zha *et al.*, “Data-centric Artificial Intelligence: A Survey,” *ACM Comput. Surv.*, vol. 57, no. 5, pp. 1–42, May 2025, doi: 10.1145/3711118.
[103] J. Jakubik, M. Vössing, N. Kühl, J. Walk, and G. Satzger, “Data-Centric Artificial Intelligence,” *Bus. Inf. Syst. Eng.*, vol. 66, no. 4, pp. 507–515, Aug. 2024, doi: 10.1007/s12599-024-00857-8.
[104] Y. Zhu, W. Moyle, M. Hong, and K. Aw, “From Sensors to Care: How Robotic Skin Is Transforming Modern Healthcare—A Mini Review,” *Sensors*, vol. 25, no. 9, p. 2895, May 2025, doi: 10.3390/s25092895.
[105] K. Liu *et al.*, “Flexible Bioinspired Healable Antibacterial Electronics for Intelligent Human-Machine Interaction Sensing,” *Adv. Sci.*, vol. 11, no. 10, p. 2305672, Mar. 2024, doi: 10.1002/advs.202305672.
[106] S. S. M. B. P. B., M. Valori, G. Legnani, and I. Fassi, “Assessing Safety in Physical Human–Robot Interaction in Industrial Settings: A Systematic Review of Contact Modelling and Impact Measuring Methods,” *Robotics*, vol. 14, no. 3, p. 27, Feb. 2025, doi: 10.3390/robotics14030027.
[107] S. Singh, M. R. Bennett, C. Chen, S. Shin, H. Ghanbari, and B. W. Nelson, “Impact of Skin Pigmentation on Pulse Oximetry Blood Oxygenation and Wearable Pulse Rate Accuracy: Systematic Review and Meta-Analysis,” *J. Med. Internet Res.*, vol. 26, p. e62769, Oct. 2024, doi: 10.2196/62769.

[108] S. Vasudevan, W. C. Vogt, S. Weininger, and T. J. Pfefer, "Melanometry for objective evaluation of skin pigmentation in pulse oximetry studies," *Commun. Med.*, vol. 4, no. 1, p. 138, Jul. 2024, doi: 10.1038/s43856-024-00550-7.

[109] Z. Tian, A. Liu, G. Zhu, and X. Chen, "A paralleled CNN and Transformer network for PPG-based cuff-less blood pressure estimation," *Biomed. Signal Process. Control*, vol. 99, p. 106741, Jan. 2025, doi: 10.1016/j.bspc.2024.106741.

[110] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045.

[111] S. Abbaspourazad, O. Elachqar, A. C. Miller, S. Emrani, U. Nallasamy, and I. Shapiro, "Large-scale Training of Foundation Models for Wearable Biosignals," 2023, *arXiv*. doi: 10.48550/ARXIV.2312.05409.

[112] X. Jing and K. Qian, "Reducing Cross-Sensor Domain Gaps in Tactile Sensing via Few-Sample-Driven Style-to-Content Unsupervised Domain Adaptation," *Sensors*, vol. 25, no. 1, p. 256, Jan. 2025, doi: 10.3390/s25010256.

[113] M. S. Lipnick, O. Ehie, E. N. Igaga, and P. Bicker, "Pulse Oximetry and Skin Pigmentation—New Guidance From the FDA," *JAMA*, vol. 333, no. 16, p. 1393, Apr. 2025, doi: 10.1001/jama.2025.1959.

[114] Y. Lin, X. Song, Y. Zhao, C. Zhang, and X. Ding, "Continuous respiratory rate monitoring through temporal fusion of ECG and PPG signals," *PLOS One*, vol. 20, no. 6, p. e0325307, Jun. 2025, doi: 10.1371/journal.pone.0325307.

[115] L. Ravaglia, M. Rusci, D. Nadalini, A. Capotondi, F. Conti, and L. Benini, "A TinyML Platform for On-Device Continual Learning With Quantized Latent Replays," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 11, no. 4, pp. 789–802, Dec. 2021, doi: 10.1109/JETCAS.2021.3121554.

[116] M. Rüb, P. Tuchel, A. Sikora, and D. Mueller-Gritschneder, "A Continual and Incremental Learning Approach for TinyML On-device Training Using Dataset Distillation and Model Size Adaption," in *2024 IEEE 7th International Conference on Industrial Cyber-Physical Systems (ICPS)*, St. Louis, MO, USA: IEEE, May 2024, pp. 1–8. doi: 10.1109/ICPS59941.2024.10639989.

[117] S. Somvanshi *et al.*, "From Tiny Machine Learning to Tiny Deep Learning: A Survey," *ACM Comput. Surv.*, vol. 58, no. 7, pp. 1–33, May 2026, doi: 10.1145/3776588.

[118] B. M. Hofstede *et al.*, "A field study to explore user experiences with socially assistive robots for older adults: emphasizing the need for more interactivity and personalisation," *Front. Robot. AI*, vol. 12, p. 1537272, Mar. 2025, doi: 10.3389/frobt.2025.1537272.

[119] A.-S. Rigaud *et al.*, "Implementation of socially assistive robots in geriatric care institutions: Healthcare professionals' perspectives and identification of facilitating factors and barriers," *J. Rehabil. Assist. Technol. Eng.*, vol. 11, p. 20556683241284765, Aug. 2024, doi: 10.1177/20556683241284765.