Università degli Studi di Cassino e del Lazio Meridionale
Corso di Dottorato in Metodi, Modelli e Tecnologie per l'Ingegneria
Dipartimento di Ingegneria Elettrica e dell'Informazione



Data-Driven Machine Learning and Deep Learning Approaches for Smart Sensor Systems and Battery State of Charge Estimation

Hamza Mustafa

hamza.mustafa@unicas.it

In Partial Fulfillment of the Requirements for the Degree of
PHILOSOPHIAE DOCTOR in
Information Engineering

21/09/2025

TUTOR
Prof. Mario Molinara
Prof. Luigi Ferrigno

COORDINATOR
Prof. Fabrizio Marignetti

UNIVERSITÀ DEGLI STUDI DI CASSINO E DEL LAZIO MERIDIONALE

CORSO DI DOTTORATO IN METODI, MODELLI E TECNOLOGIE PER L'INGEGNERIA

Date: **21/09/2025**

Author: **Hamza Mustafa**

Title: **Data-Driven Machine Learning and Deep Learning Approaches for Smart Sensor Systems and Battery State of Charge Estimation**

Department: **DIPARTIMENTO DI INGEGNERIA ELETTRICA E DELL'INFORMAZIONE**

Degree: **PHILOSOPHIAE DOCTOR**

Permission is herewith granted to university to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.



Signature of Author

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.

THE AUTHOR ATTESTS THAT PERMISSION HAS BEEN OBTAINED FOR THE USE OF ANY COPYRIGHTED MATERIAL APPEARING IN THIS THESIS (OTHER THAN BRIEF EXCERPTS REQUIRING ONLY PROPER ACKNOWLEDGEMENT IN SCHOLARLY WRITING) AND THAT ALL SUCH USE IS CLEARLY ACKNOWLEDGED.

To my family,

for your unwavering love, strength, and belief in me.

To my supervisors, Prof. Mario Molinara and Prof. Luigi Ferrigno,

for your invaluable guidance and dedication.

This thesis is a reflection of your support.

Thank you.

Acknowledgements

I would like to express my profound gratitude to my supervisors, Prof.Mario Molinara and Prof. Luigi Ferrigno, for their invaluable guidance, insightful feedback, and unwavering support throughout this doctoral journey. Their combined expertise and mentorship have been essential to the completion of this thesis. I am also grateful to Prof. Fabrizio Marignetti, coordinator of the doctoral program, for his encouragement and academic support. I wish to acknowledge the Department of Electrical and Information Engineering, whose stimulating environment has provided the resources and academic rigor necessary for this work. To my colleagues and friends, thank you for the camaraderie, thought-provoking discussions, and motivation during challenging times. A heartfelt thank you to my family for their unconditional love, patience, and belief in me. Your support has been the foundation of my perseverance. This work would not have been possible without each of you.

Contents

Acknowledgements **iv**

Acknowledgements **iv**

List of figures **ix**

List of tables **xii**

Abstract **xiv**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Application Context and Relevance 2

1.2.1 Machine Learning-Driven Spectral Color Sensing with Spectrometer Module of SENSIPATCH 2

1.2.2 Data-Driven Water Pollutant Detection Using SCW 3

1.2.3 Battery SoC Estimation Using EIS and Deep Learning 4

1.3 Research Problem and Challenges 6

1.4 Research Objectives 6

1.5 Contributions 6

1.6 Dissertation Structure 7

2 Smart Sensor Systems: SENSIPATCH and Smart Cable Water **8**

2.1 SENSIPLUS Microsensor Platform 8
2.1.1 Overview 8
2.1.2 Core Architecture 9
2.1.3 Post-CMOS Integration 9
2.1.4 Sensor Cooperation and AI Support 9
2.1.5 Distributed Sensor Deployment 9
2.1.6 Role in This Work 10
2.2 SENSIPATCH: Wearable Spectral Sensor 10
2.2.1 Hardware Configuration 10
2.2.2 Physical Design and Integration 12
2.2.3 Application to Color Sensing 12
2.3 Smart Cable Water (SCW) Sensor 13
2.3.1 Hardware Configuration 13
2.3.2 Physical Design 14
2.3.3 Usage in This Work 15
2.3.4 Applications and Capabilities 15
2.4 Data Collection Tools 16
2.4.1 SLM Studio: Software Environment for Sensor Control 16
2.4.2 Python-Based Model Development Framework 16
2.5 Summary 17

3 Related Work **18**
3.1 Related Work on Spectral Color Sensing 18
3.1.1 LED-Driven Optical Systems 18
3.1.2 Wearable and Non-Invasive Sensing 19
3.1.3 Applications in Food, Agriculture, and Environment 19
3.1.4 Advanced Imaging and Light Modeling 19
3.2 Related Work on Water Pollutant Detection 20
3.2.1 Low-Cost and IoT/WSN Systems 20

3.2.2 SENSIPLUS and Interdigitated Sensors 20
3.2.3 Machine Learning and Deep Learning Models 21
3.3 Related Work on Battery State of Charge Estimation 22
3.3.1 Traditional Model-Based and Hybrid ML Methods 22
3.3.2 Lightweight and Optimized ML Approaches 22
3.3.3 Deep Learning Architectures 23
3.3.4 EIS-Driven Approaches 23

4 Methodology **24**
4.1 Methodology for Color Sensing Using SENSIPATCH's Spectormeter Module 24
4.1.1 Preliminary Experiments: Evaluation with Limited PANTONE Colors 25
4.1.2 Extended Experiment: Classification of 100 PANTONE Colors 31
4.2 Methodology for Water Pollutant Detection Using Smart Cable Water (SCW) System 43
4.2.1 Data Collection 43
4.2.2 Data Preprocessing 46
4.2.3 Pollutant Detection Using Deep Learning Models 50

5 Results and Discussion **54**
5.1 Color Sensing: Experimental Results and Analysis 54
5.1.1 Preliminary Evaluation Using a Subset of PANTONE Colors 55
5.1.2 Comprehensive Evaluation Using 100 PANTONE Colors . . . 59
5.1.3 Discussion of Color Sensing Results 63
5.2 Water Pollutant Detection: Experimental Results and Analysis . . . 70
5.2.1 Experiment 1 Results 71
5.2.2 Experiment 2 Results 73
5.2.3 Discussion of Water Pollutant Detection Results 74

6 Battery State of Charge Estimation Using EIS and Deep Learning **78**

6.1 Dataset Generation 78
6.1.1 Battery Selection and Configuration 79
6.1.2 Data Acquisition System 79
6.1.3 Measurement Protocol 81
6.1.4 Dataset Composition and File Format 83
6.1.5 Nyquist Plot Analysis of Impedance Data 83
6.1.6 Battery SoC Estimation Using GAF-Based Image Encoding . 84
6.1.7 Transformation of EIS Data Using Gramian Angular Fields . 86
6.1.8 Deep Learning-Based Classification of GAF-Encoded EIS Data 89
6.2 Results 92
6.2.1 Overall Model Performance 92
6.2.2 Detailed Performance Metrics 93
6.2.3 Analysis of Classification Behavior 94
6.3 Summary 95

Summary and conclusions **97**

6.4 Conclusions 97
6.4.1 Spectral Color Sensing with SENSIPATCH 97
6.4.2 Water Pollutant Detection with SCW 98
6.4.3 Battery SoC Estimation with EIS 98
6.4.4 Summary of Contributions 98
6.5 Future Works 99
6.5.1 Machine Learning Methodologies 99
6.5.2 Data-Centric Improvements 100
6.5.3 Advanced Deep Learning Techniques 100
6.5.4 Dataset Expansion and Robustness 101
6.5.5 Application-Specific Extensions 101

Bibliography **102**

List of Figures

2.1 SENSIPATCH wearable device 11
2.2 Illustration of the SENSIPATCH wearable patch placed on the arm . 13
2.3 Smart Cable Water 14

4.1 Schematic Representation of the Experimental setup for preliminary data collection setup 26
4.2 ESC: Data recorded for Green Color card (PCC: 7726) under exposed conditions 29
4.3 ENC: Data recorded for Green Color card (PCC: 7726) under enclosed conditions 29
4.4 Methodology pipeline for PANTONE color sensing using data collected from the SENSIPATCH wearable system and analyzed with an MLP-based classification approach 31
4.5 Experimental Setup with Black Box Used for data acquisition 32
4.6 Schematic Representation of the data collection Setup for extended experiment 33
4.7 Spectral data captured from six PANTONE color cards under the firm contact condition (FCCW). Stable signal responses across the six sensor channels demonstrate the benefits of consistent sensor-sample proximity 37
4.8 Spectral data recorded for six PANTONE color cards under the loose contact condition (LCCW). Increased signal variability and channel overlap are evident due to the inconsistent sensor contact and possible air gaps 38
4.9 Baseline-corrected spectral data under firm contact (FCCW) for six representative PANTONE color samples. 40
4.10 Baseline-corrected spectral data under loose contact (LCCW) for six representative PANTONE color samples. 41

4.11 Graphical representation of the Raw data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs. 46
4.12 Graphical representation of the raw data, EMA, and difference values related to Acetic Acid for Platinum IDE at 78 kHz for In-phase . 48
4.13 Graphical representation of the EMA Difference data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs. 48
4.14 Graphical representation of the baseline-normalized data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs. 49
4.15 Proposed Residual Channel and Spatial Attention Network (RCSANet) 53

5.1 Stratified 5-fold cross-validation approach adopted. Each row represents the fold, and each column represents the measurement. In each fold, one measurement per class is used for testing, and the remaining four are used for training 55
5.2 Confusion Matrix for MLP model under ENC condition 58
5.3 Stratified 5-fold cross-validation across 102 classes for LCCW and FCCW data. Each row corresponds to a fold, and each column represents a measurement. In each fold, one measurement per class is used for testing, and the remaining four are used for training. . . . 60
5.4 Stratified 10-fold cross-validation across 102 classes adopted for combined data. Each row represents the fold, and each column represents the measurement. In each fold, two measurements per class are used for testing from FCCW and LCCW, and the remaining eight are used for training 64
5.5 Comparison of mean classification accuracies of machine learning models under Exposed Sensor Condition (ESC) and Enclosed Sensor Condition (ENC). 66
5.6 Classification performance of machine learning models under Firm Contact Condition With Weight (FCCW), Loose Contact Condition Without Weight (LCCW), and Combined Conditions. 67
5.7 Subset confusion matrix of the MLP model trained on combined condition data, illustrating selected PANTONE color classes with high classification accuracies 68
5.8 Subset confusion matrix for the MLP model trained on combined condition data, focusing on misclassifications among PANTONE color classes with the lowest accuracies 69

5.9 Mean Confusion matrix for CNN model 72
5.10 Mean Confusion Matrix for RCSANet 75

6.1 Schematic diagram of the implemented experimental setup. 80
6.2 Adopted procedure to obtain the experimental dataset. 82
6.3 Nyquist plots of the eleven batteries from the first test cycle. Each curve corresponds to one of the 20 acquired SoC levels. 85
6.4 Overview of the adopted methodology for SoC estimation using GAF-encoded EIS data. 86
6.5 GAF-encoded images for battery B01 at different SoC levels (after label aggregation). 88
6.6 Mean confusion matrix for the Simple CNN model across 10 folds. . 94

List of Tables

1.1 Summary of domains, sensing modalities, models, and contributions 5

2.1 SENSIPATCH sensing components, roles, and spectral characteristics 12

4.1 PANTONE Color Codes (PCC) 27
4.2 Summary of Experimental Data Collection Parameters 27
4.3 LED Features and Measurement Configurations. The columns represent Feature Name, LED Color (LC), and Wavelength (WL) in nanometers 28
4.4 Comprehensive list of PANTONE color codes used in the extended experiment. *NO_PANTONE* refers to baseline measurements without any color card placed on the sensor. *WHITE_PAPER* represents plain white paper used as a neutral reference sample 35
4.5 Summary of key parameters for data collection involving 100 PANTONE color samples under two contact conditions 36
4.6 LED feature configuration used for spectral measurements. Each entry includes the feature name, LED color, and its corresponding wavelength in nanometers 39
4.7 Model architectures and training parameters for all evaluated classifiers. 41
4.8 CNN Model Architecture 52
4.9 RCSANet Architecture 52

5.1 Classification performance for ESC 57
5.2 Classification performance for ENC 57
5.3 Mean and Standard Deviation of Classification Accuracies under ESC 59
5.4 Mean and Standard Deviation of Classification Accuracies under ENC 59
5.5 Bootstrapped dataset structure, sample distribution, and cross-validation setup 60

5.6 Classification Accuracy Under FCCW (%) 61
5.7 Mean Performance Metrics for ML Models under FCCW 62
5.8 Mean Accuracy and Standard Deviation under FCCW 62
5.9 Classification Accuracy Under LCCW (%) 62
5.10 Mean Performance Metrics for ML Models under LCCW 63
5.11 Mean Accuracy and Standard Deviation under LCCW 63
5.12 Classification Accuracy for Combined Conditions (%) 65
5.13 Mean Performance Metrics for ML Models for Combined Conditions (%) 65
5.14 Mean Accuracy and Standard Deviation under Combined Conditions 66
5.15 MLP model Class-wise accuracy across Pantone color classes for combined conditions 70
5.16 Classification Accuracy Across Ten Folds 71
5.17 Mean and Standard Deviation of Classification Accuracies 71
5.18 Classification Accuracy Across Ten Folds 74
5.19 Performance Metrics for RCSANet 76
5.20 Mean Accuracy and Standard Deviation (STD) for Each Model . . . 76

6.1 Characteristics of the adopted batteries. 79
6.2 EIS measurement file structure and column descriptions. 83
6.3 Capacity measurement file structure and column descriptions. 84
6.4 Simple CNN architecture. 90
6.5 AlexNet architecture. 91
6.6 Classification accuracy (%) of Simple CNN and AlexNet across 10-fold cross-validation. 92
6.7 Mean and Standard Deviation of Classification Accuracies for Battery SoC Estimation 93
6.8 Performance metrics (%) of Simple CNN across 10-fold cross-validation. 93

Abstract

This dissertation investigates scalable machine learning (ML) and deep learning (DL) frameworks for interpreting complex sensor signals across multiple domains. The research bridges the gap between raw signal acquisition and actionable insights by focusing on computational methods—rather than hardware innovations—to improve preprocessing, modeling, and classification of multimodal data.

Three real-world sensing applications are explored. First, wearable spectral color sensing is demonstrated using the SENSIPATCH device, which integrates LED-based spectroscopy and photodiodes for robust color classification. Through classical ML models and deep neural networks, the system achieves stable performance under variable lighting and contact conditions, with implications for biomedical, agricultural, and industrial applications. Second, water pollutant detection is addressed using the Smart Cable Water (SCW) platform, which applies electrochemical impedance spectroscopy (EIS) coupled with novel deep learning architectures, including a Residual Channel and Spatial Attention Network (RCSANet). This approach achieves over 94% accuracy across multiple pollutants, offering a scalable and cost-effective solution for real-time water quality monitoring. Third, battery state of charge (SoC) estimation is tackled using laboratory-grade EIS data, transformed into image representations via Gramian Angular Fields (GAF). Convolutional neural networks (CNNs) and AlexNet are employed to classify SoC levels, supported by a newly released public dataset from 11 lithium-iron-phosphate cells.

The thesis makes several contributions: (1) development of robust preprocessing and normalization strategies for diverse sensing modalities, (2) creation of a publicly available EIS dataset for SoC estimation, (3) validation of ML/DL models across optical, electrochemical, and impedance domains, and (4) demonstration of scalable frameworks that generalize across wearable, embedded, and laboratory platforms. Collectively, these contributions advance data-driven sensing, enabling more reliable, interpretable, and deployable solutions for environmental monitoring, energy storage, and wearable health technologies.

Keywords: Machine Learning (ML), Deep Learning (DL), Spectral Color Sensing, Wearable Sensors, Smart Cable Water (SCW), Water Pollutant Detection, Electrochemical Impedance Spectroscopy (EIS), Battery State of Charge (SoC) Estimation, Gramian Angular Field (GAF), Convolutional Neural Networks (CNNs), Attention Mechanisms (RCSANet), Environmental Monitoring, Energy Storage Systems, Data-Driven Sensing, Signal Processing.

Chapter 1

Introduction

1.1 Background and Motivation

The proliferation of smart sensors [1] and embedded systems [2] has transformed modern data-driven applications across sectors such as healthcare [3] [4], environmental monitoring [5], manufacturing [6], and smart infrastructure [7]. Smart sensors, equipped with on-board processing and communication capabilities, generate high-resolution, continuous data streams that demand robust computational methods for meaningful interpretation. However, raw sensor and signal data are often noisy, high-dimensional, and context-sensitive, making traditional signal processing methods insufficient for real-time or predictive analysis. The convergence of sensing technologies with machine learning (ML) [8] [9] and deep learning (DL) [10] [11] presents a powerful paradigm for enabling intelligent systems. These approaches facilitate automatic feature extraction, pattern recognition, and decision-making capabilities directly from complex input signals. This dissertation addresses the challenge to build scalable and effective data-driven methodologies that bridge the gap between signal acquisition and actionable insights. It focuses on the development of ML and DL frameworks tailored for diverse sensing environments, including both embedded sensor platforms and high-resolution laboratory instrumentation. Rather than emphasizing hardware innovations, the work centers on the computational and algorithmic aspects of modeling, preprocessing, and interpreting real-world signal data.

1.2 Application Context and Relevance

This research explores three real-world sensing applications that span optical, electrochemical, and impedance-based domains:

1. **Spectral color sensing** using the Spectrometer module of wearable SENSIPATCH [12] system.
2. **Chemical pollutant detection** using the Smart Cable Water (SCW) system [13].
3. **Battery State of Charge (SoC) estimation** using Electrochemical Impedance Spectroscopy (EIS) and Image-based Deep Learning.

These applications represent a range of signal types, optical reflectance spectra, electrochemical impedance curves, and frequency-domain measurements and demonstrate the versatility of machine learning methods across domains. While the first two systems are based on smart sensing platforms, the third application leverages laboratory-grade EIS data to explore SoC estimation using Gramian Angular Field (GAF) encoding and convolutional neural networks (CNNs).

1.2.1 Machine Learning-Driven Spectral Color Sensing with Spectrometer Module of SENSIPATCH

The first application domain focuses on spectral color sensing using the SENSIPATCH system, a compact, wearable device integrating a multi-wavelength LED array and photodiode-based spectrometer. Classifying colors is a fundamental task across various fields, including art [14], product development, agriculture [15], environmental assessment [16], and medical diagnostics [17]. Accurate and consistent color detection plays a crucial role across various sectors, supporting quality assurance, adherence to standards, and optimal performance. For example, in the fashion and interior design industries, maintaining precise color consistency is essential for customer satisfaction. Similarly, in the biomedical and agri-food fields, color indicators are often used to assess factors like health status or produce ripeness. Traditional color classification methods, such as visual assessments and handheld colorimeters, are often unreliable due to subjectivity, low spectral resolution, and inconsistent performance under different lighting conditions. While spectrometry provides much higher accuracy, it is typically expensive, bulky, and requires controlled lab settings. As a result, there is a rising need for compact, affordable, and real-time solutions that can function reliably in everyday environments. Innovations in discrete LED-based spectroscopy are gaining attention as a practical

alternative. These systems use fixed-wavelength LEDs to perform spectral analysis with simplified hardware, making them particularly well-suited for integration into wearable or embedded devices. One advanced example is the SENSIPATCH, a wearable system equipped with the spectrometer module, developed by Sensichips, which exemplifies the potential of this emerging technology. SENSIPATCH addresses these limitations by offering a wearable, low-cost, and real-time platform capable of accurate spectral analysis across visible and near-infrared bands. In preliminary studies, seven color samples were used to assess repeatability under exposed and enclosed lighting conditions. Subsequently, a large-scale experiment was performed with 100 color samples. A robust data processing pipeline including baseline correction, bootstrapping, and Z-score normalization was applied before evaluating classification models such as SVM, KNN, Random Forest, MLP, CNN, and LSTM. The Multilayer Perceptron (MLP) model showed excellent generalization under both firm and loose contact conditions, simulating wearable deployment scenarios. The study also opens avenues for biomedical applications such as melanoma detection and skin tone analysis. Overall, the SENSIPATCH system is demonstrated as a viable platform for wearable spectral sensing and classification.

1.2.2 Data-Driven Water Pollutant Detection Using SCW

The second application addresses chemical pollutant detection in water using the Smart Cable Water (SCW) system. This platform employs impedance spectroscopy to analyze interactions between pollutants and functionalized electrodes embedded. Water pollution represents a critical global issue with far-reaching consequences for both ecosystems and public health. The introduction of contaminants such as industrial discharges, agricultural chemicals, and domestic waste into water sources poses serious risks to aquatic organisms, vegetation, and human communities. Additionally, it significantly contributes to the widespread occurrence of waterborne illnesses, which remain among the leading causes of death worldwide [18]. Effective pollutant monitoring is essential to safeguard environmental health and ensure the sustainability of water resources. Despite progress in technological development, many water treatment facilities, especially in developing countries, continue to rely on traditional chemical analysis methods performed in laboratories. These methods are often labor-intensive, require skilled personnel, and do not support real-time monitoring, making them less suitable for tracking rapidly evolving water conditions. To address these limitations, newer monitoring strategies have begun to incorporate advanced technologies such as remote sensing (RS) and the Internet of Things (IoT). These tools enable the acquisition of large-scale, near-real-time data, offering a more accurate and responsive evaluation of water quality. However, conventional machine learning algorithms frequently face challenges when processing

the vast and complex datasets these systems produce.In contrast, DL has proven highly effective in overcoming such challenges across a range of domains, including computer vision, natural language processing, and time-series data analysis [19]. A key innovation in this field is the use of attention mechanisms, which enhance model performance by focusing on the most informative spatial and channel-specific features within the data. These mechanisms work by assigning different levels of importance to various inputs, thereby increasing both the interpretability and accuracy of the models [20]. This is particularly advantageous when working with high-noise or intricate datasets, positioning attention-based DL approaches as powerful tools in modern data analysis. The SCW system provides a real-time, IoT-compatible alternative using the SENSIPLUS [21] [22] [13] chip to capture high-dimensional, frequency-dependent electrical signals. Two major experiments were conducted. In the first, CNN-based classification was applied to distinguish eight pollutants. In the second, an expanded dataset of 14 pollutants was processed using a novel Residual Channel and Spatial Attention Network (RCSANet), incorporating attention mechanisms to enhance spatial and temporal feature extraction. This architecture achieved over 94% classification accuracy across ten-fold cross-validation, outperforming other deep learning baselines. The SCW platform offers a scalable, low-power, and cost-effective solution for real-time water quality assessment, with potential applications in industrial wastewater monitoring, aquaculture, and smart cities.

1.2.3 Battery SoC Estimation Using EIS and Deep Learning

The third application focuses on estimating battery State of Charge (SoC) using Electrochemical Impedance Spectroscopy (EIS) and deep learning-based image classification. Accurate estimation of a battery's State of Charge (SoC) is essential for the safety, efficiency, and lifespan of energy storage systems, particularly those relying on lithium-ion batteries (LiBs). In electric vehicles (EVs), SoC estimation plays a role akin to a fuel gauge, informing users of the battery's remaining usable energy. However, direct measurement of SoC is infeasible due to the nonlinear, time-varying nature of battery electrochemistry. Estimation must therefore rely on indirect measurements such as voltage, current, or impedance [23]. Conventional SoC estimation methods including ampere-hour integration [24], open-circuit voltage techniques [25], and observer-based approaches like Kalman filtering [26], sliding mode observers [27], and proportional-integral or nonlinear observers [28, 29] offer useful approximations but suffer from cumulative error propagation, sensitivity to model inaccuracies, and the need for stable operating conditions [30]. These limitations underscore the need for more adaptive and robust alternatives. Recent advancements in data-driven methods have positioned Machine Learning

(ML) and Deep Learning (DL) as promising solutions. These approaches excel in capturing nonlinear relationships within large datasets, eliminating the need for explicit model formulation. Electrochemical Impedance Spectroscopy (EIS) is particularly well-suited to this paradigm, as it provides rich frequency-domain insights into the internal electrochemical state of the battery [31]. A novel dataset [32] was generated using LFP batteries, with impedance measured at 20 SoC levels under controlled temperature and load conditions. Real and imaginary impedance values were recorded across 28 frequency points per SoC state. The EIS signals from the generated dataset were transformed into image representations using Gramian Angular Field (GAF) [33] encoding, enabling the use of convolutional neural networks (CNNs) for classification. Two DL models, Simple CNN and AlexNet were trained and evaluated across battery-wise 10-fold cross-validation.

Table 1.1: Summary of domains, sensing modalities, models, and contributions

Aspect	**Details**
Color sensing (SENSIPATCH)	
Platform	Wearable multispectral patch (LEDs + photodiode)
Modality / Data	Multispectral reflectance signals
Models	SVM, KNN, RF, MLP, CNN, LSTM
Contributions	Robust wearable spectral sensing pipeline; preprocessing and classification under variable lighting/contact
Water pollutant detection (SCW)	
Platform	Smart Cable Water (SCW) with SENSIPLUS
Modality / Data	Electrochemical impedance spectroscopy (frequency sweeps)
Models	CNN, RCSANet (residual channel & spatial attention)
Contributions	Attention-based pollutant classification (>94% accuracy, 10-fold CV); scalable real-time monitoring
Battery SoC estimation (Lab EIS)	
Platform	Laboratory EIS bench with LFP cells
Modality / Data	EIS curves converted to Gramian Angular Field (GAF) images
Models	Simple CNN, AlexNet
Contributions	Public dataset; GAF-based SoC classification; battery-wise cross-validation protocol

Table 1.1 provides a concise overview of the three investigated domains, map-

ping each sensing platform to its modality, data characteristics, applied models, and key contributions.

1.3 Research Problem and Challenges

Despite the promise of data-driven sensor and signal interpretation, significant challenges must be addressed:

- High levels of noise and variability in raw signals.
- Limited labeled datasets, especially for non-invasive or laboratory settings.
- Non-stationary behavior and signal drift across different devices or time windows.
- Trade-offs between model complexity, interpretability, and deployability.

These issues require robust preprocessing, flexible model design, domain-specific validation, and generalizable solutions.

1.4 Research Objectives

The primary objectives of this dissertation are:

- To develop complete ML/DL pipelines for real-world sensor and signal processing applications.
- To optimize preprocessing techniques for signal enhancement and feature normalization across different sensing modalities.
- To evaluate and benchmark classical ML and modern DL models for classification tasks.
- To validate the methodologies through three case studies: wearable color sensing (SENSIPATCH), chemical detection (SCW), and battery SoC estimation (EIS-based).

1.5 Contributions

The key contributions of this dissertation include:

- A two-part evaluation of the SENSIPATCH spectrometer module under controlled and scattering conditions, demonstrating robust ML-based color classification using both traditional and challenging light environments.
- A deep learning framework for pollutant detection using the SCW platform, including the novel Residual Channel and Spatial Attention Network (RCSANet).
- Generalizable preprocessing and normalization strategies applicable to both optical and electrochemical sensor data.
- A publicly available EIS dataset for SoC estimation, acquired from 11 LFP batteries under controlled experimental protocols.
- An image-based SoC classification framework using GAF encoding and deep learning (Simple CNN and AlexNet), demonstrating applicability to electrochemical signal interpretation.

1.6 Dissertation Structure

The dissertation is organized as follows:

- **Chapter 2:** Overview of sensor hardware platforms and signal modalities.
- **Chapter 3:** Review of literature on sensor data modeling and machine learning/deep learning techniques, followed by a separate review of data-driven Battery State-of-Charge (SoC) estimation methods.
- **Chapter 4:** Methodological details including experimental setup, preprocessing, model architectures, and training strategies.
- **Chapter 5:** Application-specific results, performance metrics, and comparative analysis.
- **Chapter 6:** Generation of a Battery State of Charge (SoC) dataset and its use in SoC estimation using Electrochemical Impedance Spectroscopy (EIS) and image-based deep learning models.
- **Chapter 7:** Conclusions, discussions, and directions for future work.

Chapter 2

Smart Sensor Systems: SENSIPATCH and Smart Cable Water

This chapter focuses on the hardware systems used in this research: the SENSIPLUS microsensor, the SENSIPATCH wearable system, and the Smart Cable Water (SCW) system. These hardware platforms provide the foundation for the sensor data analyzed in later chapters. While the data processing and modeling aspects are covered in the methodology chapter, here we describe the design, structure, and sensing mechanisms of the devices themselves. All systems leverage the flexible and programmable SENSIPLUS chip.

2.1 SENSIPLUS Microsensor Platform

2.1.1 Overview

SENSIPLUS is a microelectronics platform for multiple heterogeneous sensors integration into a single chip or into miniature Multi-Sensor-Microsystems (MSM). The platform is designed to address a number of environmental sensing needs, while also maintaining the scalability to support other applications.

The main challenge tackled by SENSIPLUS is the combination of heterogeneous functional materials with CMOS microelectronics into a single chip sensor a microsensor. This enables highly miniaturized, multifunctional sensing systems capable of operating in constrained environments.

2.1.2 Core Architecture

The core of the proposed system is represented by the SENSIPLUS. It is a fully programmable, single-chip integrated sensor interface that operates at low voltage and is fabricated using the UMC 0.18 μm CMOS process. Detailed information about its features can be found in [21], while here we recall the main ones.

The system includes a programmable analog front end (AFE) that is capable of both stimulating the sensor and reading its response. A Digital Sub Unit oversees system operation and facilitates communication with an external host via the Communication Line, using standard protocols such as SPI, I2C, or I3C, or alternatively through a proprietary single-wire protocol called SENSIBUS.

SENSIPLUS is capable of acquiring signals from either external sensors connected to its analog terminals (P0–P3) or from its integrated sensors (S0–S15). Additional external power pads (Pp0–Pp3) can be used to stimulate devices that require currents on the order of tens of milliamperes.

The analytical core of the chip includes a wideband Electrical Impedance Spectrometer (EIS), based on a frequency-programmable Lock-In Amplifier (LIA), and a potentiostat for electrochemical measurements. These functions are implemented within a compact 3 mm $\times$ 3 mm die, with a total power consumption of approximately 2 mW when all subsystems are active.

Enhancements to these measurement tools enable compatibility with a variety of chemical, optical, magnetic, and acoustic sensors.

2.1.3 Post-CMOS Integration

Additional integration is achieved by combining functional and sensitive materials through proprietary post-CMOS and MSM manufacturing processes. This post-fabrication step allows for the embedding of application-specific sensing elements directly onto or within the microsystem architecture.

2.1.4 Sensor Cooperation and AI Support

Multiple sensors are made to work cooperatively to improve detection and classification through the use of artificial intelligence and analytical algorithms. SENSIPLUS provides hardware-level assistance to support these computational tasks, enabling advanced interpretation of multi-sensor data.

2.1.5 Distributed Sensor Deployment

SENSIPLUS microsensors can be distributed onto Smart Cables using a proprietary single-data-wire cable bus. This system supports synchronous operation and

efficient data acquisition from distributed sensor arrays. In addition to the proprietary interface, the chip also includes standard I^2C, I^3C, and SPI serial digital interfaces, as well as analog output.

2.1.6 Role in This Work

In this dissertation, SENSIPLUS serves as the core analog-digital interface for both the SENSIPATCH and Smart Cable Water (SCW) systems. Its flexibility and multimodal capability allow seamless integration into diverse sensing configurations:

- In the **SENSIPATCH system**, SENSIPLUS controls the timing of the LED activations and synchronizes photodiode signal acquisition. It processes the optical reflectance signals using lock-in amplification techniques, enhancing signal-to-noise ratio in low-intensity conditions.
- In the **Smart Cable Water (SCW) system**, SENSIPLUS interrogates both on-chip and off-chip sensors via its high-precision Electrical Impedance Spectrometer (EIS), operating over a frequency range from 3.1 mHz to 1.2 MHz. The SCW features six Interdigitated Electrodes (IDEs), each functionalized with noble metals such as platinum and gold. The detection principle is based on redox (oxidation-reduction) reactions at the electrode surfaces, which induce frequency-dependent impedance changes.

2.2 SENSIPATCH: Wearable Spectral Sensor

2.2.1 Hardware Configuration

The SENSIPATCH, as shown in Figure 2.1, is a wearable, modular smart patch developed by Sensichips, built around the highly versatile SENSIPLUS microsensor. It integrates multiple sensing modalities within a compact, programmable platform. The full sensor suite includes:

- **Multispectral Optical Sensor:**
  - 4 visible LEDs: Blue (468 nm), Green (523 nm), Yellow (593 nm), Red (645 nm)
  - 2 near-infrared LEDs: 850 nm, 950 nm
  - 1 central photodiode with lock-in amplifier readout
- **Bioimpedance Electrode Matrix:**
  - 16 electrodes supporting bioimpedance spectroscopy (10 Hz–1.5 MHz)

  - Measurements for body composition, hydration, respiration, cardiovascular parameters

- **Gas and Environmental Sensors:**
  - Al_2O_3 ionic gas sensors for sweat/vapor detection
  - Relative humidity (RH) and temperature sensors
- **Electrochemical and Audio Sensors:**
  - Embedded potentiostat/galvanostat for voltammetry
  - Audio frequency response sensor (7 Hz–1 kHz)

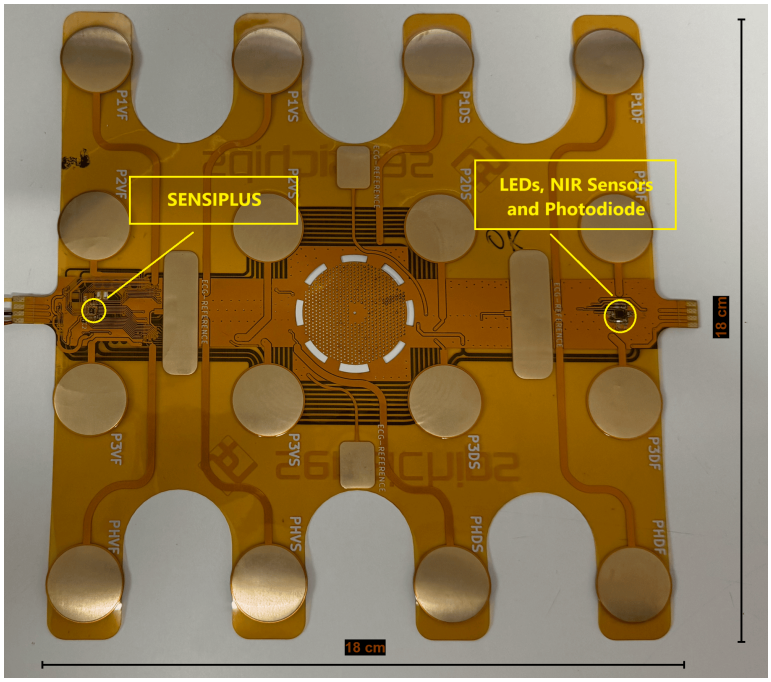


Figure 2.1: SENSIPATCH wearable device

All components are coordinated by the SENSIPLUS chip, which manages signal acquisition, analog conditioning, and synchronization.

Table 2.1 summarizes the complete sensor suite integrated within SENSIPATCH, detailing the component types, their specifications (including spectral or frequency ranges), and the corresponding functional roles.

Table 2.1: SENSIPATCH sensing components, roles, and spectral characteristics

Component Type	Specifications	Functional Role
Optical sensors		
LED array	Blue (468 nm), Green (523 nm), Yellow (593 nm), Red (645 nm); NIR: 850 nm, 950 nm	Multispectral reflectance; color discrimination
Photodiode	Lock-in amplifier readout	Signal acquisition with high SNR
Bioimpedance		
Electrode matrix	16 electrodes; 10 Hz–1.5 MHz	Body composition, hydration, respiration, cardiovascular monitoring
Gas & environmental		
Gas sensors	Al_2O_3 ionic gas detectors	Sweat/vapor detection
RH & temperature	Integrated sensor	Environmental monitoring
Electrochemical		
Potentiostat/galvanostat	Embedded in SENSIPLUS	Voltammetry; redox analysis
Audio response		
Acoustic sensor	7 Hz–1 kHz	Vibration / low-frequency response sensing

2.2.2 Physical Design and Integration

The SENSIPATCH is designed to be deployed as a wearable patch and can be integrated into wristbands, rings, necklaces, or even VR devices. Its architecture supports daisy-chaining multiple patches for distributed body sensing. For this study, the sensor was configured in a stationary experimental setup to evaluate spectral sensing using PANTONE color cards. The physical setup ensured flat and consistent contact between the sensor and color cards, isolating environmental variability. Figure 2.2 shows one of the use cases of the SENSIPATCH Wearable device.

2.2.3 Application to Color Sensing

While SENSIPATCH was not originally developed for color classification, its spectrometer module was leveraged in this work to evaluate its signal fidelity and ap-

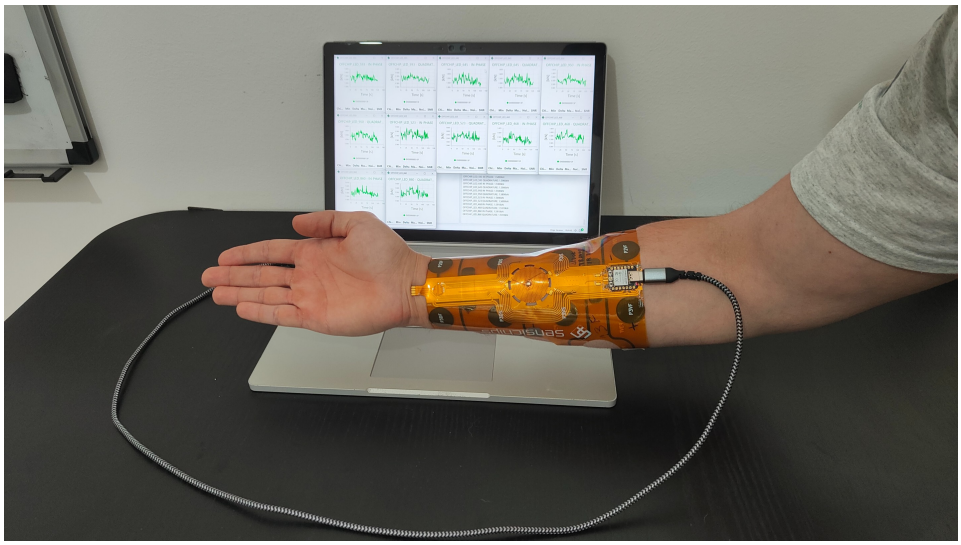


Figure 2.2: Illustration of the SENSIPATCH wearable patch placed on the arm

plicability in machine learning tasks. The goals of the experiment were:

- To benchmark signal repeatability using standardized color references (PANTONE cards)
- To develop a robust and replicable data collection methodology
- To assess multispectral sensing performance via machine learning models
- To implement and evaluate data-driven color classification using machine learning techniques

2.3 Smart Cable Water (SCW) Sensor

2.3.1 Hardware Configuration

The Smart Cable Water (SCW) sensor system, as shown in the Figure 2.3, developed by Sensichips s.r.l., is built around the **SENSIPLUS** microchip and designed for real-time water quality monitoring. The SCW integrates an advanced Electrical Impedance Spectroscopy (EIS) engine and a Potentiostat to assess a wide range of electrochemical and dielectric properties in liquids.

The sensor system supports:



Figure 2.3: Smart Cable Water

- **Electrical Impedance Spectroscopy (EIS)**: Frequency range from 3.1 mHz to 1.2 MHz, enabling detection of a wide variety of ionic and molecular species.
- **Interdigitated Electrodes (IDEs)**: Six IDEs functionalized with noble metals (Gold, Silver, Copper, Nickel, Palladium, Platinum) to enhance selectivity to different contaminants.
- **Electrochemical and Dielectric Analysis**: Includes voltammetry, conductivity, salinity, pH, redox potential, and dielectric-based turbidity and hydrocarbon sensing.
- **Environmental Monitoring**: Integrated temperature sensing with a range of -40 °C to 125 °C and ±0.1 °C accuracy.

2.3.2 Physical Design

The SCW sensor is a compact module measuring approximately 12 mm × 15 mm with a thickness of 3 mm. It includes:

- **Front-facing IDEs** for liquid contact measurements

- **Sealed back-facing Platinum IDE** for long-term stability
- **Communication interface**: I^2C and proprietary SENSIBUS for digital connectivity

Its small footprint and low power operation (0.4 mA max) make it suitable for portable or embedded water monitoring systems.

2.3.3 Usage in This Work

In this research, the SCW system was employed to collect sensor data from controlled liquid samples containing known chemical pollutants. The setup focused on:

- Capturing EIS data across a wide frequency range
- Using the functionalized IDEs to differentiate between pollutant types
- Applying machine learning models to classify pollutant types based on raw impedance data

The primary aim was to evaluate the SCW's sensing performance under laboratory conditions and to establish a robust data acquisition pipeline for pollutant classification tasks.

2.3.4 Applications and Capabilities

Beyond the scope of this study, the SCW sensor is suitable for:

- Drinking water and wastewater monitoring
- River, lake, and marine water surveillance
- Aquaculture and hydroponics system diagnostics
- Industrial effluent control
- Food and beverage process monitoring

Its broad frequency range, high selectivity due to diverse IDE materials, and integration with the SENSIPLUS platform make it a highly flexible tool for smart environmental sensing.

2.4 Data Collection Tools

Data acquisition for this research was conducted using the SENSIPLUS Learning Machine (SLM) Studio, a proprietary software platform developed by Sensichips s.r.l. SLM Studio provides a comprehensive environment for configuring, calibrating, and acquiring data from SENSIPLUS-based sensor systems, including the SENSIPATCH and Smart Cable Water (SCW) platforms. While SLM Studio supports on-device model training for rapid prototyping, the machine learning models utilized in this research were developed externally using Python-based scientific computing libraries to ensure greater methodological control, flexibility, and reproducibility.

2.4.1 SLM Studio: Software Environment for Sensor Control

SLM Studio is specifically designed to interface with SENSIPLUS-powered hardware platforms. It offers a graphical user interface for real-time sensor monitoring, experiment configuration, and labeled data collection. The software is intended for users across both technical and non-technical domains, enabling streamlined operation without requiring programming or embedded system expertise.

Key features of SLM Studio relevant to this research include:

- Real-time visualization of multimodal sensor data streams (optical, impedance, electrochemical, environmental)
- Dynamic configuration of analog front ends, excitation frequencies, LED timing, and lock-in detection parameters
- Programmable frequency sweep routines for electrical impedance spectroscopy (EIS)
- Label assignment and class-based logging, facilitating structured datasets for supervised learning
- Data export in standard machine-readable formats (CSV), supporting downstream analysis

2.4.2 Python-Based Model Development Framework

To ensure complete reproducibility and scalability of the machine learning methodology, all classification models in this research were trained using a Python-based

software stack. This approach enabled fine-grained control over feature engineering, data preprocessing, model architecture design, and evaluation protocols.

The following libraries and frameworks were employed:

- NumPy and Pandas for numerical computing, data manipulation, and statistical analysis
- Scikit-learn for classical machine learning workflows, including support vector machines, decision trees, feature scaling, and cross-validation
- TensorFlow for deep learning model construction, training, and deployment, particularly for spectral classification and impedance-based compound recognition tasks
- Matplotlib and Seaborn for data visualization and exploratory data analysis

This hybrid approach, combining sensor-native data acquisition with high-performance external modeling, ensured rigorous scientific control while leveraging the strengths of both platforms.

2.5 Summary

This chapter details the physical and functional architectures of the three core hardware components developed in this research: the SENSIPLUS microsensor chip, the spectrometer module of the SENSIPATCH wearable system, and the Smart Cable Water (SCW) electrochemical sensing system. Their shared foundation in SENSIPLUS technology enabled efficient, real-time signal acquisition and processing across very different application domains.

Chapter 3

Related Work

This chapter presents a review of existing literature relevant to the key application domains explored in this dissertation: machine learning-based spectral color sensing, deep learning approaches for chemical pollutant detection, and data-driven estimation of battery state of charge (SoC). The chapter is organized into three main sections. Section 3.1 reviews prior work on spectral color sensing, focusing on LED-based optical systems and machine learning techniques applied in wearable and embedded contexts. Section 3.2 discusses related research on water pollutant classification. Section 3.3 surveys existing approaches to battery SoC estimation.

3.1 Related Work on Spectral Color Sensing

The convergence of LED-driven spectroscopy and machine learning (ML) has fueled advancements across numerous sectors, including healthcare diagnostics, environmental surveillance, agriculture, and industrial automation. Affordable LED-based optical systems are increasingly viewed as viable substitutes for conventional spectrometers.

3.1.1 LED-Driven Optical Systems

Marsan et al.[34, 35] validated the efficacy of real-time geometrical optics for spectrophotometric analysis. Rosa et al.[36] engineered a compact optofluidic microdevice that integrates RGB LEDs with multispectral sensors for automated spectral measurements. Rocha et al.[37] used visible-light LEDs paired with ML techniques to determine liquid levels in multiphase systems. Fragner et al.[38] designed an indoor localization system using LEDs as both emitters and sensors, enhanced through ML algorithms.

3.1.2 Wearable and Non-Invasive Sensing

Wearable optical platforms are gaining traction in healthcare for non-intrusive diagnostics. Hina et al.[39] introduced a single-LED photoplethysmography system aimed at non-invasive glucose estimation. Manurung et al.[40] fused near-infrared sensing with IoT and ML to produce a portable glucose monitoring solution. Crafa et al.[41] examined the efficiency of analog vs. digital neural networks in wearable detection systems. Tobar et al.[42] developed Skinly, a portable IoT spectrophotometric device for skin diagnostics. Hsu et al.[43] created a mobile system leveraging inexpensive sensors and DL for skin reflectance measurement. Khan et al.[44] proposed an IoT-based diabetes care solution combining NIR and bio-impedance sensing with ML.

3.1.3 Applications in Food, Agriculture, and Environment

The integration of NIR/LED sensors and ML models is showing strong potential in food and agriculture. Zeb et al.[45] applied NIR LEDs with ML to identify fruit types. Sulistyo et al.[46] built a compact spectrometer to detect impurities in coconut sugar. Hanopol et al.[47] used LED-ML systems to detect rice tungro virus via reflectance patterns. Ueno et al.[48] calibrated botanical sensor arrays through supervised learning. Greco et al.[49] assessed plant hydration using terahertz imaging. Tamayo-Monsalve et al.[50] employed CNNs with transfer learning to classify coffee ripeness.

3.1.4 Advanced Imaging and Light Modeling

Goyal et al.[51] created an LED-LDR setup for chlorine and pH detection, with ML for data interpretation. Batinic et al.[52] enhanced spectral reconstruction via neural networks. Lee et al.[53] introduced a micro-LED gas sensor e-nose system powered by DL. Zhu et al.[54] combined hyperspectral imaging and ML to classify skin tones under varied lighting. Li et al.[55] used random forests for robust color classification in fashion images. Garcia et al.[56] developed an LED-driven graffiti identification system. Fan et al.[57] modeled LED spectral distributions using neural networks and genetic algorithms. Niu et al.[58] proposed VIS–NIR LED illumination for low-light scenes, while Tang et al.[59] applied hyperspectral imaging for object detection in dim environments.

These methods offer strong performance but often require complex instrumentation, limiting their use in embedded and real-time contexts. Overall, ML significantly enhances LED-based and wearable color sensing across healthcare, agriculture, and environmental domains.

3.2 Related Work on Water Pollutant Detection

Concerns over water quality and the adverse effects of pollutants on ecosystems and public health have prompted considerable progress in sensor technology. Traditional laboratory-grade sensors, while known for their high precision and reliability, are often prohibitively expensive and impractical for widespread deployment in Internet of Things (IoT)-based monitoring systems. As a result, there is growing interest in the development of low-cost sensors that strike a balance between affordability and functional adequacy for large-scale monitoring applications. Low-cost sensors have gained traction as a practical alternative for water quality monitoring, particularly in settings where resources are constrained or where extensive spatial coverage is required. Their cost-effectiveness makes them accessible to a wide range of users, including municipalities, industries, and citizen scientists. For example, carbon-based sensors originally intended for air quality applications have shown potential for adaptation to water quality monitoring due to their low production costs and ease of deployment [60]. Despite these advantages, low-cost sensors often face challenges related to calibration, accuracy, and reliability. However, advancements in sensing technologies and data processing techniques have significantly improved their performance. Singh et al. [61] demonstrated that such sensors can reliably capture real-time water quality parameters, offering valuable insights for water resource management and decision-making.

3.2.1 Low-Cost and IoT/WSN Systems

The integration of IoT technologies with low-cost sensors has further transformed the field of environmental monitoring. IoT-enabled devices allow real-time data transmission to cloud platforms, facilitating continuous observation and early detection of pollution events. Wireless Sensor Networks (WSNs) have proven effective in tracking critical water quality parameters such as pH, conductivity, turbidity, and dissolved oxygen, thus supporting sustainable water management strategies [62, 63]. Moreover, when coupled with cloud computing and advanced analytics, these systems can identify pollution patterns and predict contamination risks [64]. This fusion enhances responsiveness and enables proactive interventions in water management systems.

3.2.2 SENSIPLUS and Interdigitated Sensors

To overcome the limitations of basic low-cost sensors, researchers have begun exploring advanced sensor technologies such as interdigitated sensors. For instance, the SENSIPLUS platform developed by Sensichips offers a compact, sensitive, and

selective solution for detecting a wide range of pollutants. These sensors can be seamlessly integrated into IoT infrastructures for real-time analysis. Liu et al. [65] showed that combining interdigitated sensors with deep learning algorithms significantly improves pollutant classification and detection accuracy. Several recent studies have proposed systems that monitor both air and water quality using platforms like SENSIPLUS [66, 67, 68, 69, 70, 71]. These systems support a variety of monitoring tasks—from binary contaminant detection to more advanced pollutant classification—typically through customized, application-specific deployments.

3.2.3 Machine Learning and Deep Learning Models

In parallel, machine learning (ML) techniques have gained prominence in water quality assessment and classification tasks. Nasir et al. [72] evaluated a range of ML models, including Support Vector Machines (SVM) and Random Forest, to classify water quality using the Water Quality Index (WQI). Similarly, [73] used Random Forest classifiers to predict and categorize water quality levels. Dilmi et al. [74] applied both Long Short-Term Memory (LSTM) networks and SVM to classify water into three categories using physicochemical data. Decision tree-based approaches have also been explored. For instance, [75] developed a decision tree model to predict six key water quality indicators, including pH, temperature, chemical oxygen demand (COD), ammonia-nitrogen (NH_3-N), and nitrate-nitrogen (NO_3-N). Haghiabi et al. [76] compared the performance of SVM, Artificial Neural Networks (ANN), and the Group Method of Data Handling (GMDH) in estimating the water quality of Iran's Tireh River. While these studies have contributed significantly to water quality monitoring, most focus on general classification rather than the identification of specific pollutants.

Only a limited number of works combine sensor-based data acquisition with advanced ML or deep learning models for pollutant-specific classification. Alzubi et al. [77] presented an IoT-based monitoring system using KNN and SVM to assess general water pollution indicators such as pH, temperature, conductivity, and turbidity, without addressing specific contaminants. Konde et al. [78] designed a Water Quality Monitoring system using FPGA and XBee communication to monitor physical parameters like turbidity and CO_2, but it lacks pollutant-specific detection. Zhu et al. [79] employed ConvLSTM networks to analyze water pollution based on nitrate, dissolved oxygen, biological oxygen demand, total coliform, and fecal coliform, though their focus remained narrow. Similarly, Kang et al. [80] used LSTM models to interpret nematode behavior in detecting formaldehyde and benzene, without addressing a broader range of pollutants. Some recent efforts aim to advance pollutant-specific classification. For example, [71] proposed a low-cost sensor platform using natural language generation to identify wastewater pollu-

tants, achieving improved accuracy. Charulatha et al. [81] used principal component regression and ANN to estimate nitrate content in groundwater, while [82] applied partial least square discriminant analysis to identify explosive compounds in sewage. Dean et al. [83] used CNN and LSTM networks to classify chemicals in seawater. Banerjee et al. [84] proposed an ML model based on static environmental factors to predict key water quality parameters, but their work did not address pollutant classification directly. Ferdinandi et al. [85] developed an ANN model with a single hidden layer to classify five pollutants using data collected via the SENSIPLUS chip. In a subsequent work [86], Gerevini et al. incorporated KNN, SVM, and anomaly detection techniques for improved identification of pollutants in water samples.

3.3 Related Work on Battery State of Charge Estimation

Accurate SoC estimation is critical for lithium-ion batteries in EVs and energy storage. Traditional methods (OCV, Coulomb counting, Kalman filters) suffer from cumulative errors and sensitivity to conditions. Data-driven ML/DL approaches provide adaptive, robust alternatives.

3.3.1 Traditional Model-Based and Hybrid ML Methods

Hybrid models combining neural networks with Kalman filtering dominate this category. Likun et al.[87] used BPNN with DEKF; Pang et al.[88] combined BPNN with EKPF; Lian et al.[89] fused BPNN with UKF and RLS. Jiashun et al.[90] applied Adaboost-BPNN–UKF, while Wang et al.[91] optimized anti-interference tracking. These approaches improved accuracy and robustness over traditional baselines.

3.3.2 Lightweight and Optimized ML Approaches

Extreme Learning Machines (ELMs) and kernel methods offer lightweight yet effective solutions. Molla et al.[92] combined ELM with GSA; Cheng et al.[93] proposed adaptive sequence ELM; Dou et al.[94] used MI-ELM with RLS; Zhao et al.[95] hybridized ELM with SCASSA optimization; Carlos et al.[96] applied ELM+PSO. Kernel-based methods include SVR+PSO [97], LS-SVM+UPF [98], LS-SVM+GWO [99], and online LS-SVM+AUKF [100].

3.3.3 Deep Learning Architectures

DL models are powerful for capturing nonlinear battery dynamics. Junxiong et al. [101] proposed EI-LSTM-CO; Paul et al. [102] developed RLSTM with attention; Yong et al. [103] introduced HIPF filtering; Li et al. [104] combined fractional models with immune algorithms. Wang et al. [105] proposed robust ELM with correntropy; Hu et al. [106] used a DNN with Coulomb counting. CNN-LSTM hybrids also show promise: Bian et al.[107], Yang et al.[108], Chen et al.[109], and Zou et al.[110] achieved high accuracy. Random forest-based models such as Lipu et al.[111] also report competitive results.

3.3.4 EIS-Driven Approaches

Electrochemical Impedance Spectroscopy (EIS) provides detailed frequency-domain information. Singh et al.[112] integrated EIS with fuzzy logic; Xu et al.[113] applied fractional-order modeling; Westerhoff et al.[114] combined EIS with ECMs. Messing et al.[115] trained dual DNNs; Babaeuyazdi et al.[116] compared GPR vs LR; Buchicchio et al.[117] achieved 93% accuracy with compact classifiers; Bourelly et al.[118] optimized EIS measurement time while preserving accuracy. Together, these studies demonstrate the potential of EIS-driven ML pipelines for SoC estimation.

Chapter 4

Methodology

This chapter presents the data-driven methodologies developed for the design, implementation, and optimization of smart sensor systems, with a particular focus on classification tasks. The work was carried out in collaboration with SENSICHIPS Srl, whose advanced sensing technologies formed the foundation of the experimental platforms utilized in this research. Emphasis is placed on the integration of SENSICHIPS sensors with deep learning and machine learning models to enable accurate classification across different application domains. The methodological framework encompasses the complete experimental pipeline, including sensor system design, data acquisition, signal preprocessing, and model development. Two primary smart sensor platforms are investigated in this study, for each system, the specific experimental procedures, data-driven strategies, and classification models are described.

4.1 Methodology for Color Sensing Using SENSIPATCH's Spectormeter Module

The PANTONE color cards were used as standardized physical samples for collecting spectral data across various color categories. In each measurement session, a color card was placed directly on the surface of the SENSIPATCH device to ensure uniform and reproducible exposure to the sensing system, which comprises a spectrometer module featuring multi-wavelength LEDs, near-infrared (NIR) sensors, and a photodiode. These components operate together to illuminate the sample and capture reflected light across multiple wavelengths, enabling detailed spectral analysis of each color. The spectral data acquisition was managed using Sensichips' SLM Studio software, as described in chapter 2 which controls the sensor system

and records the photodiode responses for each light source. By serving as consistent and industry-standard references, the PANTONE [119] cards allowed for reliable comparison across sessions and supported accurate evaluation of the SENSIPATCH system's color recognition performance. The collected spectral data were then used to train and evaluate machine learning models for color classification. The study was conducted in two phases: an initial pilot experiment using a small subset of PANTONE colors to validate the methodology, followed by a comprehensive experiment involving 100 color samples under standardized and controlled conditions.

The methodological framework common to both experiments is structured into three principal phases:

1. **Data Collection**
2. **Data Preprocessing**
3. **Classification Using Machine Learning Models**

Each of these phases is described in detail in the subsequent sections. This structured approach ensures consistency, reproducibility, and a systematic evaluation of the SENSIPATCH system's performance across different experimental scales.

4.1.1 Preliminary Experiments: Evaluation with Limited PANTONE Colors

The first series of experiments aimed to assess the feasibility and initial performance of the SENSIPATCH system. These preliminary tests involved a reduced set of PANTONE color samples and two distinct environmental conditions.

Experimental Setup:

Before initiating data collection, it was essential to establish a controlled environment to obtain consistent and reliable spectral readings from the PANTONE color samples. These color samples, recognized as standardized hues widely used across industries to maintain color accuracy, served as an ideal reference point for evaluating the SENSIPATCH sensor system. Two distinct experimental conditions were employed:

- Exposed Sensor Condition (ESC): The sensors were initially uncovered to record ambient light conditions. Afterward, a PANTONE color card was placed directly on top of the sensors.

- Enclosed Sensor Condition (ENC): A black enclosure was placed over the sensors at the beginning of the trial. Once the baseline data was recorded, the enclosure was briefly lifted to position a PANTONE color card onto the sensors and then replaced.

This methodology was designed to assess sensor performance under both standard and fluctuating lighting environments. An abstract schematic of the experimental setup used in the preliminary data collection is shown in Figure 4.1.

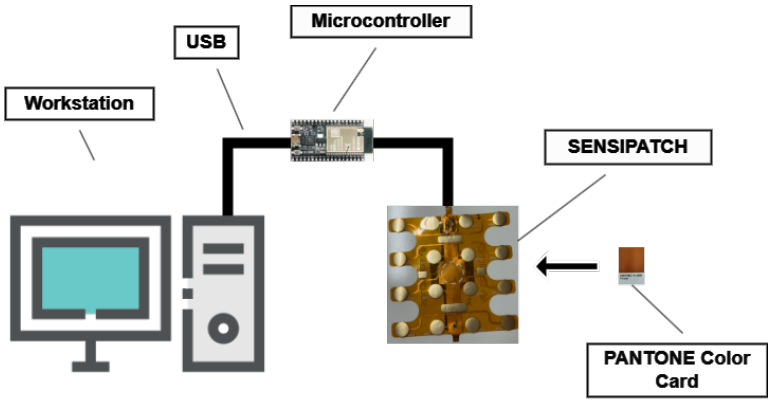


Figure 4.1: Schematic Representation of the Experimental setup for preliminary data collection setup

Data Collection:

For each of the eight sample types (seven PANTONE colors and one Air reference), five independent measurement sessions were conducted under both the Exposed Sensor Condition (ESC) and the Enclosed Sensor Condition (ENC), with the color cards placed directly on the sensors during each session. Data collection involved standardized handling of the cards to ensure consistent contact with the sensor surface and eliminate variability due to physical alignment or distance. The PANTONE Color Codes (PCC) corresponding to each color card used in the experiments are listed in Table 4.1.

At the start of each measurement session, data was recorded for 150 samples without any color card placed on the sensor surface, capturing the baseline "Air" condition. Subsequently, a PANTONE color card was positioned directly on the sensor, ensuring zero-gap contact for consistent spectral acquisition. Data collection then continued until the 1151st sample. This procedure was repeated in

Table 4.1: PANTONE Color Codes (PCC)

Color	PCC
White (Lightest Sky)	11-4804
Blue (Caribbean Sea)	18-4525
Black (Pirate Black)	19-4305
Green	7726
Yellow	7548
Purple	5125
Red	1797

five independent sessions for each of the eight sample types (seven PANTONE colors and Air) under both experimental conditions (ESC and ENC). The SENSIPATCH system integrates four visible-wavelength LEDs and two near-infrared (NIR) sources, which are activated sequentially. A single photodiode captures the reflected signal from each light source, resulting in six measurements per acquisition cycle. Each light source requires approximately 265 milliseconds for readout, so a complete cycle takes around 1.6 seconds, yielding an effective sampling rate of 0.62 full cycles per second. Each session lasted approximately 30 minutes.

A summary of the experimental parameters and data acquisition specifications is provided in Table 4.2.

Table 4.2: Summary of Experimental Data Collection Parameters

Parameter	Value/Description
Number of Sample Types	8 (7 PANTONE colors + Air)
Measurement Sessions per Sample Type	5
Samples per Session	1151
Baseline Samples (Air) per Session	150
Color Card Samples per Session	1001
Total Number of Samples	46,040
Number of Light Sources	6 (4 LEDs + 2 NIR)
Photodiode Read Time per Light Source	265 ms
Time for One Full Acquisition Cycle	1.6 seconds
Sampling Rate per Full Cycle	0.62 Hz (readings/sec)
Duration per Session	30 minutes

Figures 4.2 and 4.3 illustrate the data collected for the green color card (PCC: 7726) under ESC and ENC conditions, respectively. As shown in these figures, ESC data displays more variation due to ambient light interference, whereas ENC data remains stable after the color sample PANTONE card is placed, emphasizing

the benefit of controlled lighting conditions. For the purpose of training machine learning models, the raw measurements from the four visible LEDs and two near-infrared (NIR) sources were directly used as features, resulting in six time series inputs. Each feature corresponds to the in-phase current response of the photodiode to a specific LED stimulus. Feature names encode the measurement configuration, including the sensor output port, applied bias settings, and the activated LED wavelength. For example, the feature `OUTPORT_PORT_SDO_DCBIASP_200_IN_PHASE` represents the in-phase response recorded using the yellow LED (593 nm) via the PORT SDO output, with a positive DC bias of 200 mV. This systematic naming convention is applied to all sensor features, reflecting the distinct combinations of LED wavelengths and biasing conditions used during data acquisition. Details of these features and associated configurations are summarized in Table 4.6.

Table 4.3: LED Features and Measurement Configurations. The columns represent Feature Name, LED Color (LC), and Wavelength (WL) in nanometers

Feature Name	**LC**	**WL**
OUTPORT PORT SDO DCBIASP 200 IN-PHASE	Yellow	593
OUTPORT PORT INT DCBIASP 70 IN-PHASE	Red	645
OUTPORT AUX DCBIASP 35 IN-PHASE	Infrared	950
OUTPORT SHA DCBIASP 25 IN-PHASE	Infrared	850
OUTPORT PORT RDY DCBIASP 35 IN-PHASE	Green	523
OUTPORT PORT CS DCBIASP 25 IN-PHASE	Blue	468

The SENSIPATCH device operates as a compact spectrometer with an integrated lock-in amplifier capable of isolating both the in-phase (IN-PHASE) and quadrature (QUADRATURE) components of a sinusoidal signal. In this application, the sensors are stimulated using constant current sources, and the photodiode responses measured as current represent the reflected light intensity. Since the photodiode response is a real signal, analysis focused on the IN-PHASE component, which provides a direct and proportional measure of the incident light-induced current. This component was therefore selected as the key input feature for downstream analysis.

Preprocessing

Before training the Machine Learning (ML) models, the collected sensor data underwent a preprocessing step to remove systematic bias. Specifically, a baseline adjustment was applied by subtracting the mean of the first 150 baseline samples recorded when no PANTONE color card was present and the sensor was exposed only to ambient environmental light from all subsequent measurements. This step

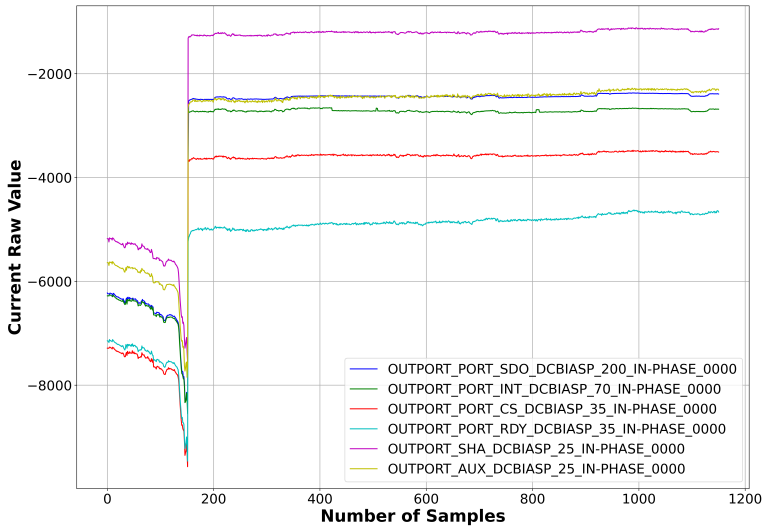


Figure 4.2: ESC: Data recorded for Green Color card (PCC: 7726) under exposed conditions

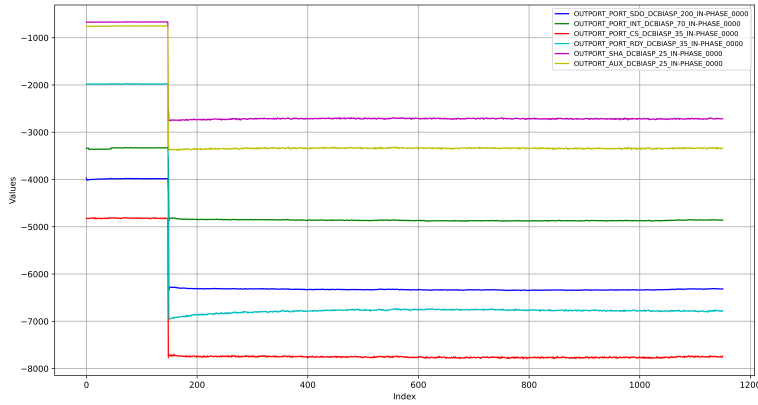


Figure 4.3: ENC: Data recorded for Green Color card (PCC: 7726) under enclosed conditions

helped eliminate the influence of ambient light and sensor offset, ensuring the resulting values reflected only the sensor's interaction with the color stimuli.

The correction is defined in Equation 4.1:

$$X_{\text{unbiased}} = X - \text{mean}(X_{\text{baseline}}) \tag{4.1}$$

Here, X_{unbiased} is the corrected sensor data, X is the raw measurement, and $\text{mean}(X_{\text{baseline}})$ is the mean of the baseline samples.

Colors Classification

After the data acquisition phase, the next stage involved classifying the collected color data using machine learning (ML) techniques. Several widely used and well-established ML models were evaluated for this purpose, including the Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). These models were chosen based on their strong track records in classification tasks and their complementary strengths in dealing with different types of feature distributions and decision boundaries.

- **Multi-Layer Perceptron (MLP)**: Selected due to its capacity to model complex, non-linear relationships between input features, making it especially suitable for high-dimensional color spectrum classification.
- **K-Nearest Neighbors (KNN)**: Applied for its simplicity and intuitive logic, relying on the similarity between feature vectors in Euclidean space to infer color categories.
- **Random Forest (RF)**: Utilized for its ensemble learning capability, combining multiple decision trees to improve prediction accuracy and reduce overfitting.
- **Support Vector Machine (SVM)**: Employed for its effectiveness in non-linear classification through kernel-based decision boundaries, which enable it to perform well in high-dimensional feature spaces.

Among the evaluated models, the MLP achieved the highest classification accuracy. To fine-tune its performance for the PANTONE color recognition task, a set of model hyperparameters was empirically selected. The MLP architecture consisted of a single hidden layer, with the number of neurons set to three times the number of input features. Given that the input vector comprises six features (reflectance values from four visible LEDs and two NIR sources), the hidden layer was configured with 18 neurons. This design choice aimed to balance the model's

representational capacity with the complexity of the input data, ensuring sufficient flexibility to capture intricate spectral patterns without leading to overfitting.

Additional hyperparameters were also optimized to improve training efficiency and model stability. A learning rate of 0.5 and a momentum value of 0.1 were adopted based on empirical validation. To assess model generalization and avoid bias due to random training/testing splits, a 5-fold cross-validation strategy was implemented. In this approach, the dataset was partitioned into five equal parts; during each iteration, one part was used for validation while the remaining four were used for training. This rotation ensured that all data points contributed to both the training and evaluation processes across different folds, resulting in a more robust and generalizable model assessment.

4.1.2 Extended Experiment: Classification of 100 PANTONE Colors

To evaluate the capabilities of the SENSIPATCH system in handling complex color classification tasks, an extended series of experiments was conducted using 100 distinct PANTONE color card samples. This diverse set of physical samples was selected to cover a wide range of color tones and spectral characteristics, allowing for a comprehensive assessment of the system's scalability, sensitivity to subtle color differences, and overall classification performance. This phase employed a revised experimental setup, incorporating controlled conditions and additional evaluation strategies to more rigorously test the system's performance. Figure 4.4 shows the full sensor-to-classification pipeline for color sensing using the SENSIPATCH.

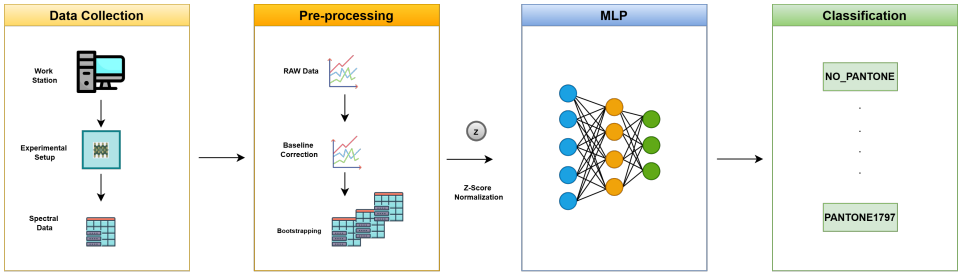


Figure 4.4: Methodology pipeline for PANTONE color sensing using data collected from the SENSIPATCH wearable system and analyzed with an MLP-based classification approach

Experimental Setup

A dedicated experimental setup was developed to ensure the consistent and accurate acquisition of spectral data while minimizing the influence of ambient lighting. To achieve this, measurements were conducted within a custom-built black-box enclosure, internally lined with black paper sheets. Figure 4.5 represents the experimental setup used for data acquisition. This controlled environment was specifically designed to eliminate external light interference, which is known to affect the reliability of spectral measurements. The SENSIPATCH wearable sensor system was positioned inside this enclosure in direct contact with the PANTONE color cards samples, simulating realistic usage scenarios where the device would typically operate in close proximity to the skin or another surface. This configuration was intended to maintain consistent sensor alignment and contact pressure, thereby enhancing the reproducibility of readings and suppressing external optical noise. A schematic illustration of the complete experimental setup is shown in Figure 4.6.

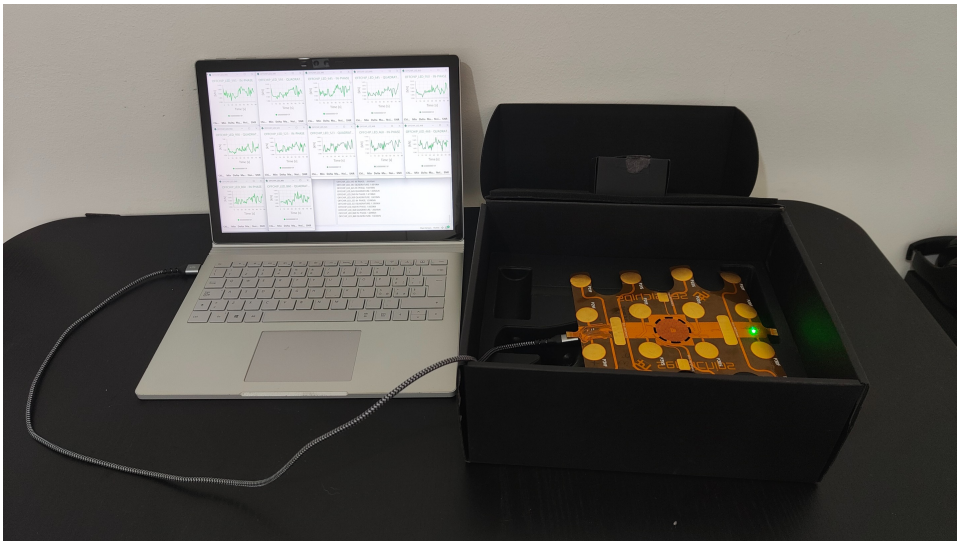


Figure 4.5: Experimental Setup with Black Box Used for data acquisition

Data Collection Under Varying Contact Conditions To investigate how physical contact pressure influences the quality of spectral measurements, data collection was performed under two distinct sensor-to-sample interface conditions:

- **Firm Contact Condition (FCCW) – With Applied Weight:** A custom-designed circular weight with a hollow center was placed on top of the PAN-

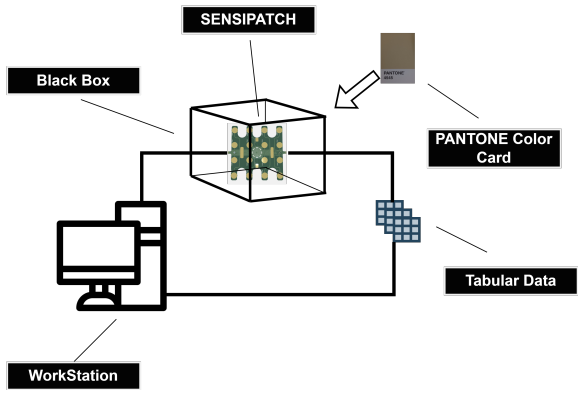


Figure 4.6: Schematic Representation of the data collection Setup for extended experiment

TONE color card to maintain consistent, evenly distributed pressure between the sensor and the sample. The hollow design ensures that the sensor's optical path remains unobstructed, allowing full light emission and detection. This condition mimics scenarios in wearable sensing where tight contact enhances signal consistency by reducing air gaps.

- **Loose Contact Condition (LCCW) – Without Applied Weight:** In this configuration, the color card was gently laid over the sensor without any additional force. This setup represents real-world usage cases in which wearable sensors may not always maintain perfect contact, such as during motion or temporary detachment from the skin.

These contrasting conditions were introduced to evaluate the sensitivity of the spectral measurements to contact variability. Since proximity and pressure can significantly influence reflectance behavior, this comparison helps assess the robustness of the sensing system under practical conditions.

Measurement Procedure A systematic data acquisition process was implemented to ensure consistency and repeatability across all measurement sessions. The procedure included the following steps for each contact condition:

- **Sensor Placement:** The SENSIPATCH device was positioned inside the controlled black-box enclosure, and the lid was closed to eliminate interference from ambient lighting.

- **Baseline Signal Recording:** To capture the intrinsic sensor signal and background light response, the first 21 measurements were recorded without placing any external object on the sensor. These baseline readings were used to account for sensor noise and drift, a known issue in wearable devices.
- **Color Card Placement:** After baseline recording, a PANTONE color card was applied to the sensor surface under the assigned contact condition. The enclosure was sealed again to maintain consistent lighting.
- **Spectral Measurement Collection:** The system then recorded measurements 22 through 121, capturing the sensor's response to the specific color sample.
- **Full Dataset Generation:** This process was repeated for all 100 PANTONE color cards, generating a large and balanced dataset suitable for statistical and machine learning analysis.

The controlled conditions ensured that any observed variation in the spectral data could be attributed to the sensor–sample interaction, rather than uncontrolled environmental influences. The use of a black-box enclosure and careful control over contact mechanics were critical in achieving reliable and reproducible measurements.

Data Collection

To construct a comprehensive dataset for evaluating the SENSIPATCH wearable system's ability to classify PANTONE colors, spectral data were acquired from 100 individual PANTONE color cards, along with an additional plain white paper sample used as a reference. The full list of PANTONE color codes used in this study is provided in Table 4.4. The data acquisition protocol was carefully designed to maintain consistency, reduce variability, and ensure reliability across repeated measurements.

Each color sample underwent five separate trials, with each trial capturing 121 sensor readings. During measurements, color cards were placed directly on the sensor surface to ensure uniform contact and eliminate any air gaps, which is critical for minimizing inconsistencies caused by varying sensor-to-sample distances.

For each trial, the first 21 readings were taken without a color card on the sensor which served as the baseline segment. Following the baseline, the respective PANTONE color card was positioned on the sensor, and the remaining 100 readings were collected to capture its spectral profile under controlled conditions. This approach allowed the system to detect any drift or shift in sensor behavior and correct it through baseline normalization.

Table 4.4: Comprehensive list of PANTONE color codes used in the extended experiment. *NO_PANTONE* refers to baseline measurements without any color card placed on the sensor. *WHITE_PAPER* represents plain white paper used as a neutral reference sample

No.	PANTONE Code	No.	PANTONE Code	No.	PANTONE Code	No.	PANTONE Code
1	NO_PANTONE*	2	PANTONE100	3	PANTONE108	4	PANTONE11-4804
5	PANTONE12-0313	6	PANTONE12-0524	7	PANTONE12-0727	8	PANTONE12-5204
9	PANTONE13-0645	10	PANTONE13-0850	11	PANTONE13-1107	12	PANTONE13-5304
13	PANTONE14-0848	14	PANTONE14-0852	15	PANTONE14-1119	16	PANTONE14-1419
17	PANTONE14-1513	18	PANTONE14-1911	19	PANTONE14-3612	20	PANTONE14-4516
21	PANTONE142	22	PANTONE15-0543	23	PANTONE15-0942	24	PANTONE15-1247
25	PANTONE15-1816	26	PANTONE15-2706	27	PANTONE15-5210	28	PANTONE15-6437
29	PANTONE152	30	PANTONE16-1255	31	PANTONE16-1329	32	PANTONE16-1432
33	PANTONE16-1439	34	PANTONE166	35	PANTONE17-1210	36	PANTONE17-1436
37	PANTONE17-1456	38	PANTONE17-1564	39	PANTONE17-4021	40	PANTONE17-4524
41	PANTONE17-5722	42	PANTONE1797	43	PANTONE18-0119	44	PANTONE18-1248
45	PANTONE18-2326	46	PANTONE18-3025	47	PANTONE18-4330	48	PANTONE18-4525
49	PANTONE18-4530	50	PANTONE18-5621	51	PANTONE19-1250	52	PANTONE19-2025
53	PANTONE19-3438	54	PANTONE19-3921	55	PANTONE19-3935	56	PANTONE19-4305
57	PANTONE19-6110	58	PANTONE199	59	PANTONE214	60	PANTONE277
61	PANTONE291	62	PANTONE3278	63	PANTONE345	64	PANTONE350
65	PANTONE383	66	PANTONE394	67	PANTONE3985	68	PANTONE414
69	PANTONE4545	70	PANTONE466	71	PANTONE4685	72	PANTONE476
73	PANTONE485	74	PANTONE493	75	PANTONE5125	76	PANTONE535
77	PANTONE5405	78	PANTONE5555	79	PANTONE583	80	PANTONE600
81	PANTONE605	82	PANTONE635	83	PANTONE656	84	PANTONE677
85	PANTONE691	86	PANTONE732	87	PANTONE7415	88	PANTONE7493
89	PANTONE7548	90	PANTONE7590	91	PANTONE7654	92	PANTONE7720
93	PANTONE7726	94	PANTONE7737	95	PANTONE9041	96	PANTONE9044
97	PANTONE9180	98	PANTONE9241	99	PANTONE9320	100	PANTONECoolGray7
101	PANTONEWarmGray2	102	WHITE_PAPER*				

For each measurement sequence, the sensors were first exposed to ambient conditions—without any PANTONE color card—to establish a baseline response. After recording this baseline, a color card was placed on the sensors, and data collection proceeded for 121 consecutive readings. This approach was intended to capture any potential sensor drift over time, thereby improving the accuracy and reliability of the dataset used for machine learning classification.

In total, measurements were collected for 100 different PANTONE color cards, a white paper reference, and a baseline condition representing the absence of any material on the sensors. Under two experimental conditions, and with five repeated trials per card at 121 readings per trial, the resulting dataset contained 123,420 samples (61,170 per condition; calculated as $102 \times 5 \times 121 \times 2$).

The data acquisition rate was optimized to maintain efficiency while preserving accuracy. Each photodiode measurement was taken roughly every 265 milliseconds, with all six sensors recorded sequentially, resulting in an overall sampling interval of about 1.6 seconds per complete sensor cycle.

The specific LED configurations, wavelengths, and feature names for the data collection for extended experimentation are summarized in Table 4.6.

Figures 4.7 and 4.8 visualize the raw sensor outputs for six selected PANTONE colors under the firm and loose contact conditions, respectively. As evident from the plots, firm contact leads to smoother and more stable signals across all six sensor channels. In contrast, loose contact introduces greater variability, likely caused by inconsistent surface proximity and minor air gaps. These observations highlight the significance of maintaining proper sensor contact in wearable spectral sensing scenarios.

An overview of the data acquisition parameters used for the full experiment is provided in Table 4.5.

Table 4.5: Summary of key parameters for data collection involving 100 PANTONE color samples under two contact conditions

Parameter	**Value/Description**
Number of PANTONE Colors	100
Additional Sample	Plain White Paper
Trials per Color	5
Sensor Readings per Trial	121
Baseline Samples per Trial	21
Experimental Conditions	2 (Firm Contact – FCCW, Loose Contact – LCCW)
Total Samples	121,000 ($100 \times 5 \times 121 \times 2$)
Photodiode Read Time per Light Source	265 ms
Light Sources Used	6 (4 LEDs + 2 NIR)
Time per Full Acquisition Cycle	1.6 seconds
Sampling Rate	0.62 full cycles/sec
Sensor Output Components	In-Phase (used), Quadrature (available)
LED Configuration	See Table 4.6
Contact Conditions	FCCW (firm), LCCW (loose)

Data Preprocessing

To ensure that the spectral data collected were suitable for machine learning-based classification, a multi-step preprocessing pipeline was implemented. This process aimed to enhance the consistency, accuracy, and reliability of the dataset. The preprocessing involved three primary operations: baseline correction, dataset augmentation through bootstrapping, and feature normalization using Z-score standardization. These steps were carried out after data acquisition to refine the signals recorded from the 100 PANTONE color samples and to improve the generalization capability of the classification models.

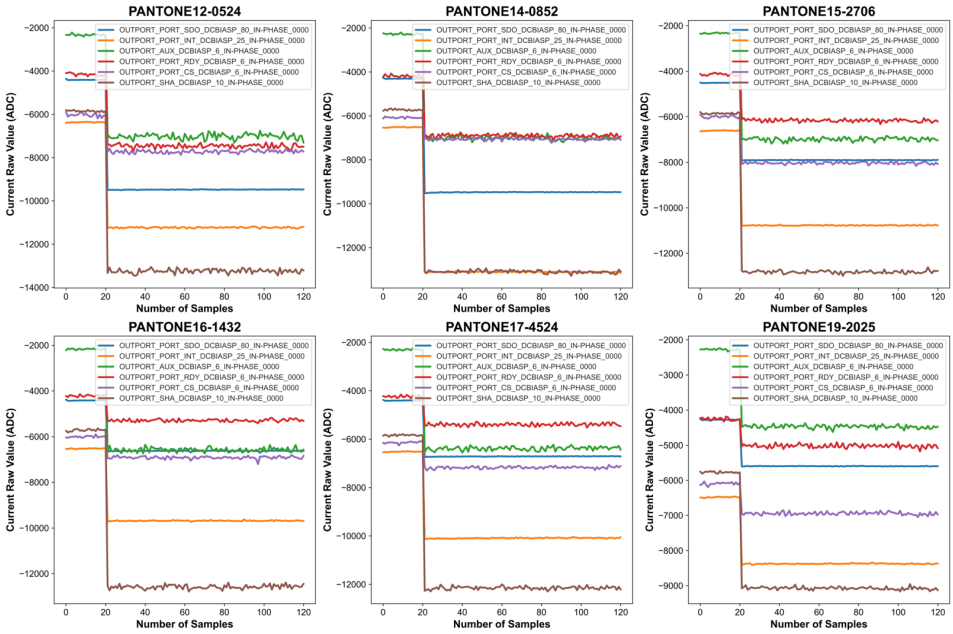


Figure 4.7: Spectral data captured from six PANTONE color cards under the firm contact condition (FCCW). Stable signal responses across the six sensor channels demonstrate the benefits of consistent sensor-sample proximity

Baseline Correction

Prior to applying any machine learning techniques, the raw spectral measurements were adjusted to account for the inherent sensor signal in the absence of any external stimulus. This was achieved by implementing a baseline correction step, which used the initial 21 readings from each trial captured without any PANTONE card on the sensor surface as the baseline reference. These values reflect the background response of the system and are used to correct the subsequent measurements.

As seen in Figures 4.7 and 4.8, there is a noticeable jump in signal intensity when the color card is introduced. To correct for this, the mean of the first 21 baseline readings was calculated and subtracted from each of the remaining 100 readings taken with the card in place.

The baseline adjustment was mathematically formulated as:

$$X_{\text{corrected},i} = X_{\text{raw},i} - X_{\text{baseline_mean}}, \quad \forall i \in \{22, 23, \ldots, 121\} \tag{4.2}$$

where:

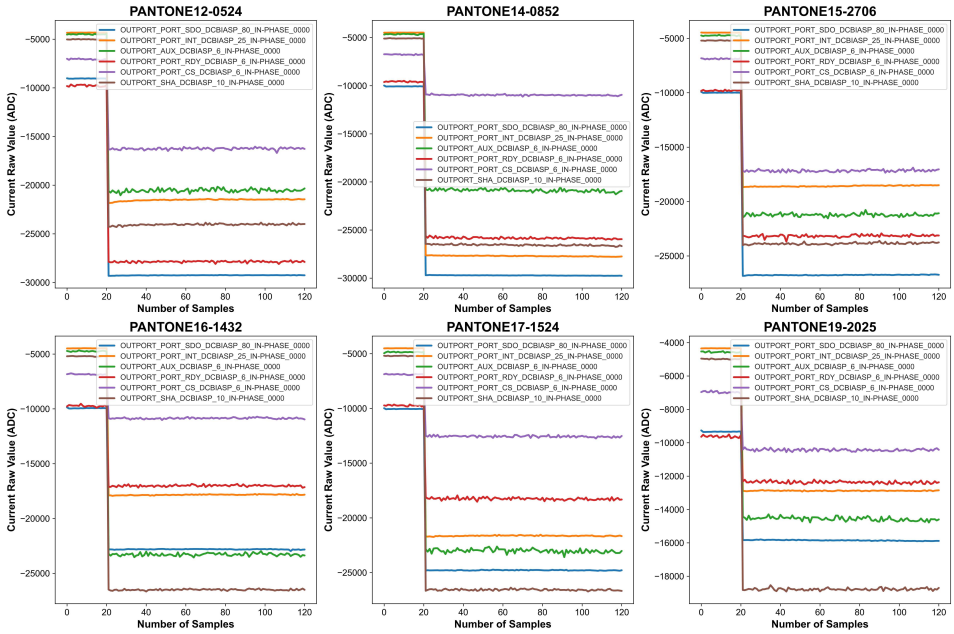


Figure 4.8: Spectral data recorded for six PANTONE color cards under the loose contact condition (LCCW). Increased signal variability and channel overlap are evident due to the inconsistent sensor contact and possible air gaps

- $X_{\text{raw},i}$ represents the original unprocessed measurement at time step i,
- $X_{\text{baseline_mean}}$ is the mean of the 21 baseline values,
- $X_{\text{corrected},i}$ denotes the resulting baseline-adjusted reading.

This operation aligned all spectral measurements to a common zero-reference point, thereby eliminating sensor drift and environmental bias. The effect of this correction is visually demonstrated in Figures 4.9 and 4.10, where the adjusted signals show better separation between channels and reduced noise, particularly under firm contact conditions.

Bootstrapping for Dataset Augmentation

To strengthen the dataset and reduce the likelihood of overfitting, a bootstrapping approach was used to artificially expand the number of training samples. For each trial file containing 100 baseline-corrected readings, a total of 2000 samples were generated through random sampling with replacement.

Table 4.6: LED feature configuration used for spectral measurements. Each entry includes the feature name, LED color, and its corresponding wavelength in nanometers

Feature Name	LC	WL
OUTPORT PORT SDO DCBIASP 80 IN-PHASE	Yellow	593
OUTPORT PORT INT DCBIASP 25 IN-PHASE	Red	645
OUTPORT AUX DCBIASP 6 IN-PHASE	Infrared	950
OUTPORT SHA DCBIASP 10 IN-PHASE	Infrared	850
OUTPORT PORT RDY DCBIASP 6 IN-PHASE	Green	523
OUTPORT PORT CS DCBIASP 6 IN-PHASE	Blue	468

This method preserved the original statistical characteristics of the dataset while introducing controlled variability, making the model more robust to noise and slight inconsistencies in the data. The increased number of training instances also improved the stability and accuracy of the learning algorithms by exposing them to a broader range of input variations.

Feature Normalization Using Z-score Standardization

Following bootstrapping, Z-score normalization was applied to standardize all input features across the six spectral channels. This transformation adjusted each feature to have a mean of zero and a standard deviation of one, ensuring uniform scaling across all sensor outputs.

The normalization equation is expressed as:

$$X_{\text{normalized}} = \frac{X_{\text{bootstrapped}} - \mu}{\sigma} \tag{4.3}$$

where:

- $X_{\text{bootstrapped}}$ is the spectral reading after bootstrapping and baseline correction,
- μ is the mean of the resampled values,
- σ is the standard deviation of the resampled values.

Because each LED-photodiode pair operates at a unique wavelength and exhibits varying intensity levels, normalization was crucial to neutralize these differences. This allowed the machine learning models to focus on relative spectral patterns rather than absolute amplitude values, which improved classification accuracy and consistency across different color tones.

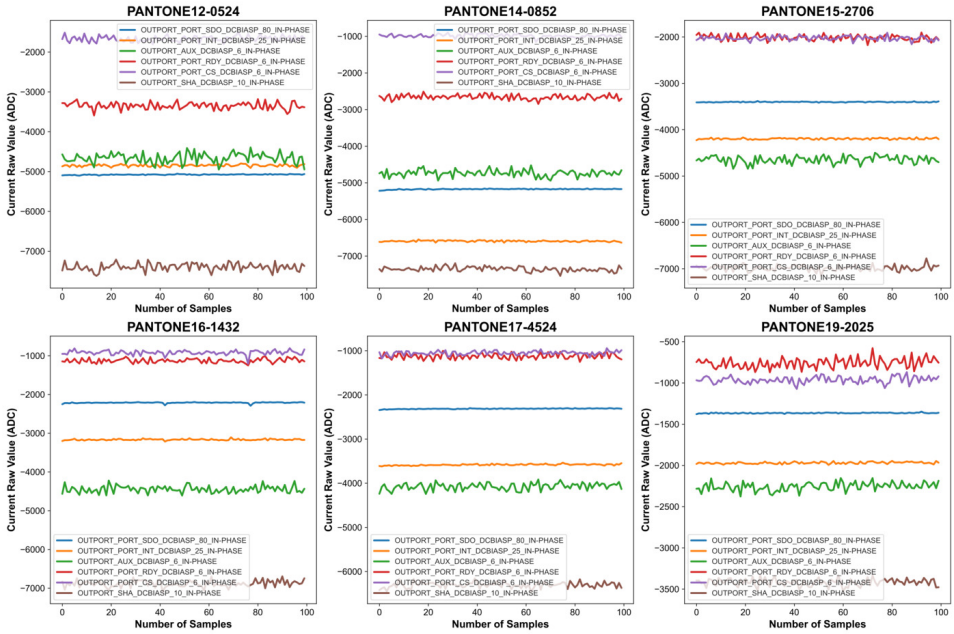


Figure 4.9: Baseline-corrected spectral data under firm contact (FCCW) for six representative PANTONE color samples.

Color Classification with Machine Learning

Following the data preprocessing stage, machine learning techniques were applied to classify PANTONE color samples using spectral data collected from the SENSIPATCH wearable system. The objective was to build a classification framework capable of distinguishing between a wide range of colors with high accuracy, while maintaining generalizability across varied experimental settings. For the input for the machine learning models, each sample comprises six raw features—the photodiode's in-phase current responses to sequential LED activations (four visible, two near-infrared).

Model Selection and Rationale

A combination of traditional and deep learning models was evaluated to explore different strategies for spectral pattern recognition. Three models, Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF), were retained from the pilot experiments due to their strong baseline performance and interpretability. In the extended experiments, two deep learning architec-

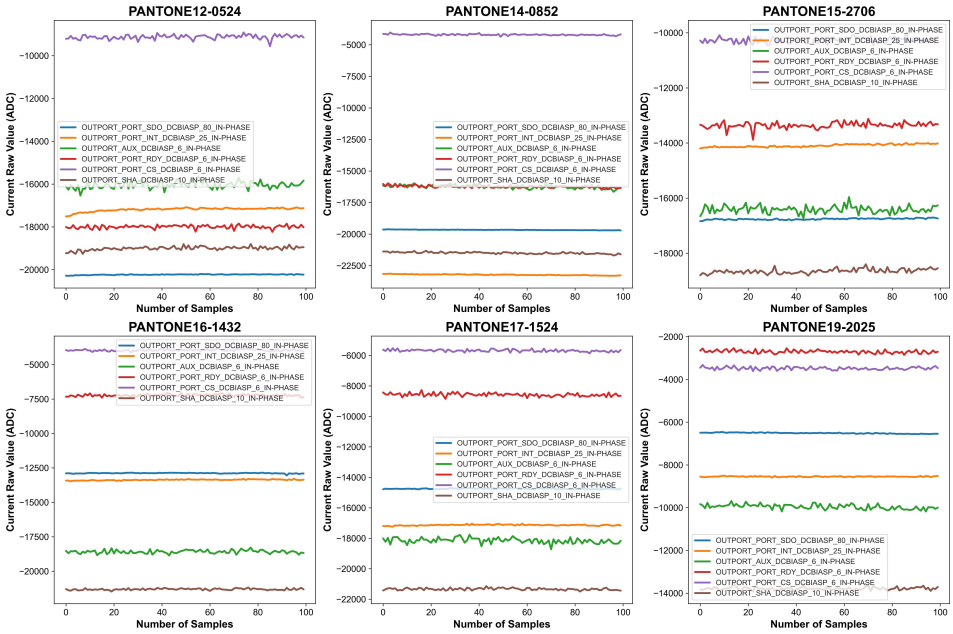


Figure 4.10: Baseline-corrected spectral data under loose contact (LCCW) for six representative PANTONE color samples.

tures—Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—were introduced to investigate their effectiveness in modeling more complex spectral structures.

Parameter	MLP	Random Forest	SVM	2D CNN	LSTM
Key Layers	Dense (100, 50)	–	–	Conv2D (16,32), Dense (64)	LSTM (32,32), Dense (32)
Activation Function	ReLU	–	–	ReLU	ReLU (Dense)
Optimizer	Adam	–	–	Adam	Adam
Learning Rate	–	–	–	0.001	0.001
Regularization	$\alpha = 0.001$	–	–	–	–
Loss Function	Categorical Cross-Entropy	–	–	Sparse Categorical Cross-Entropy	Sparse Categorical Cross-Entropy
Batch Size	64	–	–	32	16
Epochs	50	–	–	20	20
Window Size	–	–	–	5	5
Additional Notes	MaxIter=500	200 Estimators	Linear Kernel, Prob=True	Dropout(0.5), BatchNorm	Dropout(0.5), BatchNorm

Table 4.7: Model architectures and training parameters for all evaluated classifiers.

- **Random Forest (RF)**: Included for its ensemble learning capabilities and strong performance in high-dimensional feature spaces. RF also enables feature importance analysis, providing interpretability alongside robustness.
- **Support Vector Machine (SVM)**: Selected for its proficiency in handling complex decision boundaries, especially valuable when spectral classes exhibit

non-linear separability.

- **Multilayer Perceptron (MLP)**: Chosen as a strong baseline model due to its proven ability to learn non-linear mappings in spectral data. The MLP's dense architecture is particularly effective for structured numerical inputs like LED response data.
- **Convolutional Neural Network (CNN)**: Introduced in the extended phase to explore local feature extraction from spectral inputs. Despite their origin in image processing, CNNs have shown success in spectral classification by identifying region-specific patterns through convolutional filters.
- **Long Short-Term Memory (LSTM)**: Also added in the extended study to capture sequential dependencies between wavelength channels. LSTM networks, as a form of recurrent neural network, are well-suited to model temporal or ordered data such as spectral sequences.

Among all models, the MLP consistently offered the best generalization across both contact conditions. SVM achieved the highest accuracy in the firm contact setting, while RF performed well across the full dataset. However, these traditional models showed reduced flexibility compared to the MLP when conditions varied. Table 4.7 outlines the configurations and hyperparameters used for each classifier.

Implementation of Multilayer Perceptron (MLP)

To capitalize on its flexibility and effective modeling of non-linear relationships, the MLP architecture was designed with multiple hidden layers and trained using cross-validation. The input consisted of six spectral features, corresponding to LED responses. A variety of architectures were tested, including networks with one to three hidden layers. Each configuration was trained using the Adam optimizer and categorical cross-entropy loss function.

The MLP's forward propagation can be defined as:

$$\mathbf{h}^{(0)} = \mathbf{x} \tag{4.4}$$

$$\mathbf{h}^{(l)} = \phi\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right), \quad \text{for } l = 1, 2, \ldots, L-1 \tag{4.5}$$

$$\hat{\mathbf{y}} = \text{softmax}\left(\mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}\right) \tag{4.6}$$

Here, ϕ represents the ReLU activation function, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the layer weights and biases, and $\hat{\mathbf{y}}$ is the predicted class probability vector over the PANTONE categories.

The training objective was to minimize the categorical cross-entropy loss:

$$\mathcal{L}_{\text{CCE}} = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}) \tag{4.7}$$

Where y_{ij} denotes the ground truth label and $\hat{y}_{ij}$ is the predicted probability for the jth class of the ith sample.

This layered architecture allows the MLP to learn complex patterns in spectral input data. The use of ReLU activations promotes non-linearity, while the softmax output enables probabilistic classification. Cross-validation further strengthens model generalization by evaluating performance across multiple data splits.

4.2 Methodology for Water Pollutant Detection Using Smart Cable Water (SCW) System

This section outlines the methodology employed for detecting and classifying chemical pollutants in water using the Smart Cable Water (SCW) system. Two sets of experiments were conducted to explore the design and optimization of a data-driven pipeline combining advanced sensing hardware and deep learning models. Both experiments were based on the SENSIPLUS impedance spectroscopy platform, with distinct differences in preprocessing techniques, pollutant sets, and learning architectures. The methodological framework is common to both experiments and is structured into three phases:

1. **Data Collection**
2. **Data Preprocessing**
3. **Pollutant Detection Using Deep Learning Models**

Each of these phases is described in the following subsections. This structured approach ensures consistency, reproducibility, and systematic evaluation of the Smart Cable Water for pollutant detection.

4.2.1 Data Collection

The data collection procedure was standardized across both experiments to ensure measurement consistency and reproducibility. The sensing platform used in this study is the Smart Cable Water (SCW) system. While full technical specifications and sensor configurations are provided in the introductory material, this section focuses on the aspects of the hardware setup that directly influenced data acquisition.

The SCW sensor employs six Interdigitated Electrodes (IDEs), each composed of a different conductive material: gold, silver, platinum, nickel, palladium, and copper. These electrodes generate frequency-dependent impedance signatures in response to ionic and molecular interactions in water. For each measurement, both in-phase and quadrature components of the impedance signal were captured over a defined frequency range, yielding a complex spectral profile for every sample.

Each chemical class was evaluated independently under controlled laboratory conditions using the following acquisition steps:

1. **Baseline Stabilization:** An initial 600 impedance readings were recorded in clean, deionized water to establish the system's baseline response. Measurements were acquired at a frequency of 0.5 Hz to ensure stability.

2. **Pollutant Introduction:** The target pollutant was gradually added to the water till it was fully dissolved or added to the water.

3. **Pollution Monitoring Phase:** Following pollutant injection, the system recorded an additional 1,000 impedance samples under the same acquisition settings.

This protocol was applied uniformly across all chemical classes investigated in both experiments, resulting in a total of 1,600 impedance samples per pollutant. The consistency in measurement conditions and sample volume was maintained to minimize experimental variability and ensure fair model evaluation in later classification stages.

Pollutant Sample Sets

Two separate experiments were designed to evaluate the SCW system's ability to discriminate among a variety of waterborne chemical pollutants based on their impedance signatures.

Experiment 1 – Initial Evaluation (8 Pollutants):

- Acetic Acid, Ammonia, Hydrogen Peroxide, Hydrochloric Acid,
- Phosphoric Acid, Sodium Chloride, Sodium Hydroxide, Sodium Hypochlorite.

These compounds were chosen to represent a wide range of physicochemical properties, including variations in ionic strength, acidity/alkalinity, and redox behavior, thus allowing a robust assessment of sensor sensitivity and specificity.

Experiment 2 – Extended Evaluation (14 Pollutants):

In the second experiment, the same eight pollutants were retained, and six additional substances were introduced to increase chemical diversity and better simulate real-world water contamination scenarios:

- Acetone, Calcium Nitrate, Ethanol, Formic Acid, Nelsen, Potassium Nitrate.

These additions included both organic and inorganic pollutants, enhancing the complexity of the classification task and enabling a broader test of the system's generalization capabilities.

Measurement Channels Selected from Impedance Data

Each trial produced raw impedance signals saved in CSV format via SLM Studio. From these files, a subset of ten measurement channels was selected for analysis based on signal reliability and their consistent response across test conditions. These channels correspond to combinations of electrode type, signal component (in-phase or quadrature), and frequency.

Measurements were recorded from four off-chip Interdigitated Electrodes (IDEs): gold, platinum, silver, and nickel. Two frequencies, 200Hz and 78kHz, were selected to capture distinct electrochemical interaction profiles. Low-frequency signals emphasize diffusion and bulk properties, while high-frequency measurements are more sensitive to electrode-surface interactions and double-layer effects.

The ten selected measurement channels were:

1. OFFCHIP_PLATINUM_78kHz_IN-PHASE
2. OFFCHIP_GOLD_78kHz_IN-PHASE
3. OFFCHIP_PLATINUM_200Hz_IN-PHASE
4. OFFCHIP_PLATINUM_200Hz_QUADRATURE
5. OFFCHIP_GOLD_200Hz_IN-PHASE
6. OFFCHIP_GOLD_200Hz_QUADRATURE
7. OFFCHIP_SILVER_200Hz_IN-PHASE
8. OFFCHIP_SILVER_200Hz_QUADRATURE
9. OFFCHIP_NICKEL_200Hz_IN-PHASE
10. OFFCHIP_NICKEL_200Hz_QUADRATURE

Each of these channels consists of a time series of 1,600 samples per trial, capturing the full temporal evolution from baseline stabilization to pollutant exposure. Figure 4.11 represent the Raw data collected with SCW for five pollutants. These raw outputs were then used as the input for the preprocessing and classification stages described in the following sections.

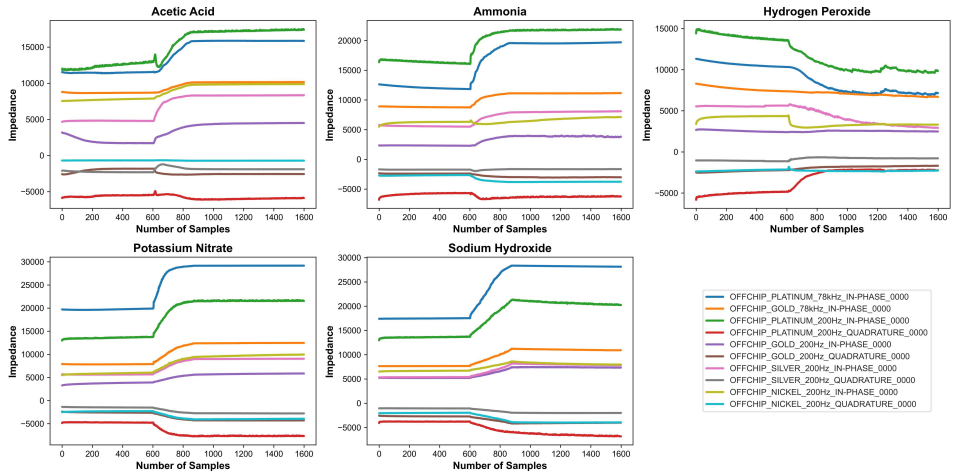


Figure 4.11: Graphical representation of the Raw data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs.

4.2.2 Data Preprocessing

To ensure that the raw impedance signals collected from the SCW system were suitable for deep learning-based classification, a structured preprocessing pipeline was applied. The preprocessing steps were tailored separately for each experiment to account for differences in data characteristics and sensor behavior. In Experiment 1, a noise-suppressing Exponential Moving Average (EMA) followed by differencing was used to highlight signal deviations. In Experiment 2, baseline normalization was performed using an exponential smoothing filter initialized on clean water samples. In both cases, a fixed-length sliding window was applied to convert the continuous multivariate time series into overlapping segments suitable for input into neural network models. The following sections detail the normalization procedures and windowing strategy used in each experiment.

Experiment 1: EMA-Difference Normalization

The initial stage of data preprocessing in Experiment 1 involved smoothing the raw impedance signals using an Exponential Moving Average (EMA) filter. This technique reduces noise and highlights underlying signal trends. The EMA for each data point was calculated using Equation 4.8:

$$E_{EMA_i} = \left(\frac{a-1}{a}\right) E_{EMA_{i-1}} + \left(\frac{1}{a}\right) x_{EMA_i} \tag{4.8}$$

Once the EMA curve was computed, the next step involved calculating the deviation of the raw data from the smoothed trend. This was done by subtracting the EMA values from the original signal (OD), producing what is referred to as the EMA Difference (EMA-D). This transformation, defined in Equation 4.9, enhances local fluctuations that may correspond to meaningful signal changes useful for classification:

$$EMA\text{-}D_i = OD_i - E_{EMA_i} \tag{4.9}$$

Following this transformation, the first 400 samples of each recording were discarded. This was done to remove transient behavior at the beginning of each measurement, which could be affected by sensor stabilization or initial chemical interactions. Retaining only the steady-state portion of the signal reduces noise and improves the relevance of the input to the learning model. The remaining EMA-D sequence was segmented using a sliding window approach. Figure 4.12 shows an example of the raw impedance measurements, the EMA smoothed curve, and the resulting EMA-D signal for the Acetic Acid sample, recorded using the off-chip platinum IDE at 78 kHz (in-phase component). Figure 4.13 shows the EMA-D signals for five pollutants, recorded using the off-chip platinum IDE at 78 kHz (in-phase component).

Experiment 2: Baseline Normalization

In alignment with the methodology established in Experiment 1, the preprocessing phase for Experiment 2 was designed to prepare impedance time-series data for input into deep learning models. The same ten impedance signal channels were selected as input features, comprising in-phase and quadrature measurements at 78 kHz and 200 Hz from four off-chip IDEs (platinum, gold, silver, and nickel). These are listed in Section 4.2.2 and are reused here to maintain consistency across both experiments. Unlike the EMA-difference technique applied in the first experiment, the preprocessing in Experiment 2 employs a baseline normalization strategy. This approach leverages the initial 600 samples collected from clean, potable water in

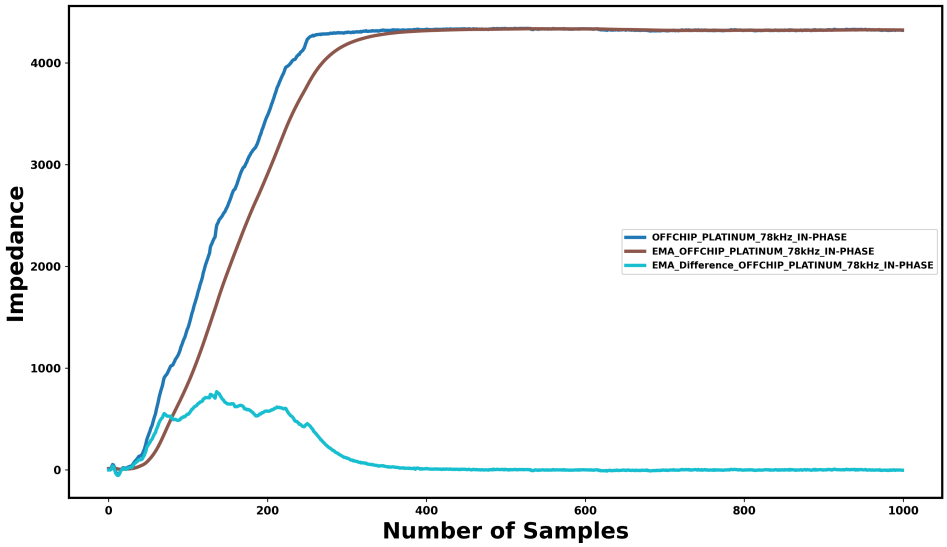


Figure 4.12: Graphical representation of the raw data, EMA, and difference values related to Acetic Acid for Platinum IDE at 78 kHz for In-phase

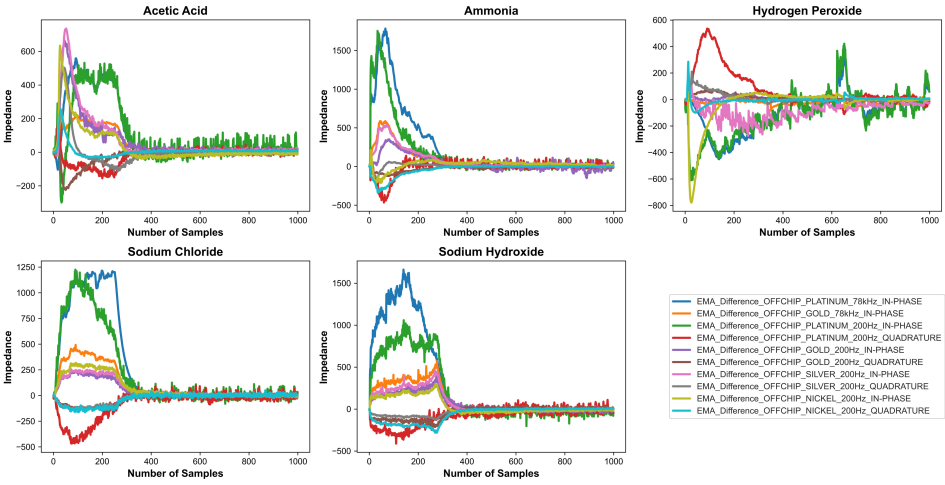


Figure 4.13: Graphical representation of the EMA Difference data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs.

each measurement to establish a dynamic reference baseline for subsequent pollutant exposure readings. Normalization was performed using an exponentially weighted moving average, defined by the recurrence relation in Equation 4.10:

$$b_t = \begin{cases} s_t & \text{if } t = 0, \\ \lambda s_t + (1-\lambda) b_{t-1} & \text{if } t > 0 \end{cases} \tag{4.10}$$

where $\lambda = \frac{1}{25}$, s_t is the raw sensor signal at time t, and b_t is the smoothed baseline value at the same time step.

Once baseline normalization was applied, 1,000 time samples were retained from each pollutant class for model training and evaluation. This standardization process minimized signal drift and ensured that the subsequent analysis focused on chemically induced impedance changes rather than artifacts from sensor initialization or environmental variation. Figure 4.14 illustrates the baseline-normalized signals profile for five pollutants, displaying representative impedance responses across both frequency domains and signal components. This normalization technique enhances inter-class comparability and improves signal quality for data-driven modeling.

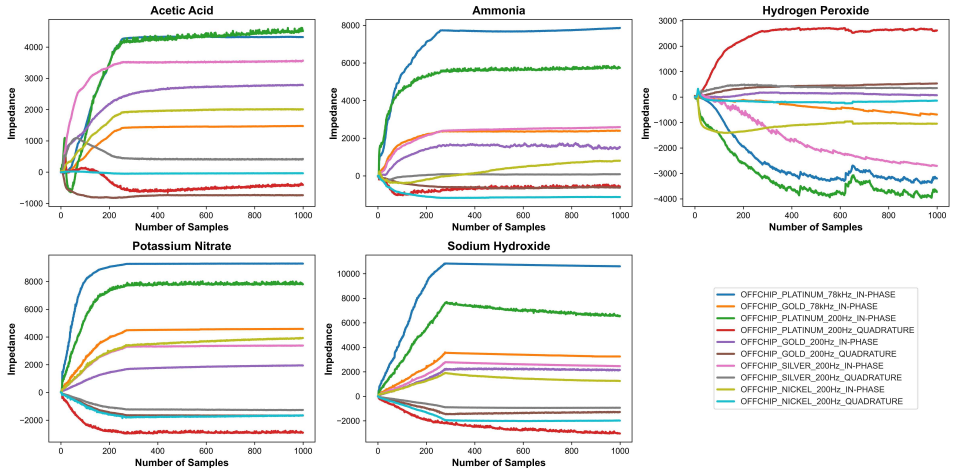


Figure 4.14: Graphical representation of the baseline-normalized data for five pollutants, showing In-phase and Quadrature components at 78 kHz and 200 kHz from six IDEs.

4.2.3 Pollutant Detection Using Deep Learning Models

Following the data preprocessing steps described in the previous subsection, a series of deep learning models were employed to classify pollutants based on the processed impedance time-series data. The selected models were designed to capture spatial and temporal dependencies inherent in the multichannel sensor signals. For Experiment 1, a baseline evaluation was conducted using four widely recognized architectures: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM) models. These models were selected for their established effectiveness in sequence modeling and pattern recognition tasks. Among them, the CNN model demonstrated the highest classification performance and is therefore discussed in detail in the following section. In Experiment 2, the same baseline models were reused to ensure consistency in comparative evaluation. To further enhance classification accuracy, two additional deep architectures were introduced: a Residual Neural Network (ResNet) and a novel model proposed in this study, Residual Channel and Spatial Attention Network (RCSANet). RCSANet outperformed all other models in Experiment 2 and is presented as the principal contribution of this phase. The following subsections focus on the CNN and RCSANet architectures, as they exhibited the best performance in their respective experimental settings.

Sliding Window Segmentation

Prior to training, the impedance time-series data were segmented using a sliding window technique to construct fixed-length input samples for all deep learning models used in both experiments. This approach enables the models to capture short-term temporal dependencies and localized patterns that are critical for classifying pollutant-specific responses. Let N denote the total number of data points in a single sequence and F the number of input features. A window of size W is moved across the sequence with a stride S_t, producing $N - W + 1$ overlapping segments. Each window $X_i \in \mathbb{R}^{W \times F}$ represents a local temporal context, defined as:

$$X_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,F} \\ x_{i+1,1} & x_{i+1,2} & \cdots & x_{i+1,F} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i+W-1,1} & x_{i+W-1,2} & \cdots & x_{i+W-1,F} \end{bmatrix}$$

where $x_{i,j}$ denotes the jth feature at the ith time step. For all experiments in this study, the window size was set to $W = 32$ and the stride $S_t = 1$, providing fine-grained segmentation with maximum temporal overlap. This segmentation

step produces structured input samples that are passed to the subsequent models. It is particularly well-suited for architectures such as CNNs, LSTMs, and attention-based models, which rely on temporal structure to learn discriminative representations from sequential data.

Convolutional Neural Network (CNN)

Among the models evaluated in Experiment 1, the Convolutional Neural Network (CNN) architecture demonstrated the highest classification performance. CNNs are particularly effective in extracting localized patterns from structured input data, making them well-suited for handling the windowed time-series segments generated from the impedance signals.

Although traditionally used in image classification tasks, CNNs can be effectively applied to multivariate time-series data by treating each sliding window as a two-dimensional input matrix. In this context, each input sample has dimensions $W \times F$, where $W = 32$ is the window length and $F = 10$ represents the number of impedance-based features.

The CNN model processes this input using a series of convolutional and pooling layers that progressively extract hierarchical temporal features. These are followed by fully connected (dense) layers that map the extracted features to one of the eight pollutant classes. The model uses ReLU activation functions throughout the hidden layers and a softmax layer for multi-class classification at the output.

The overall CNN architecture was designed to balance expressive capacity with computational efficiency, enabling robust learning while avoiding overfitting. Table 4.8 summarizes the specific layers, activation functions, and training hyperparameters used in the implementation.

Residual Channel and Spatial Attention Network (RCSANet)

For Experiment 2, we evaluated several deep learning models for classifying 14 pollutant types, including CNN, LSTM, RNN, and ResNet architectures. Among these, the proposed Residual Channel and Spatial Attention Network (RCSANet) achieved the highest performance. RCSANet extends the ResNet backbone by incorporating attention mechanisms that enhance the model's ability to selectively focus on informative features both across channels and within temporal windows. The complete architectural flow of RCSANet is illustrated in Figure 4.15, and its configuration is detailed in Table 4.9.

Table 4.8: CNN Model Architecture

Layer Type	Configuration
Input	Shape: (32, 10, 1)
Conv2D	Filters: 32, Kernel: (3, 3), Activation: ReLU
Batch Normalization	-
Conv2D	Filters: 64, Kernel: (3, 3), Activation: ReLU
Batch Normalization	-
Dropout	Rate: 0.5
MaxPooling2D	Pool size: (1, 1)
Conv2D	Filters: 128, Kernel: (3, 3), Activation: ReLU
Batch Normalization	-
Flatten	-
Dense	Units: 128, Activation: ReLU
Dense	Units: 64, Activation: ReLU
Output (Dense)	Units: Number of classes, Activation: Softmax

Table 4.9: RCSANet Architecture

Layer Type	Configuration
Input	Input shape: (32, 10, 1)
Convolutional	64 filters, kernel size: (3, 3), activation: ReLU
Batch Normalization	–
Max Pooling	Pool size: (2, 2)
Residual Block 1	64 filters
Channel Attention	Reduction ratio: 16
Spatial Attention	Kernel size: (7, 7), activation: Sigmoid
Residual Block 2	128 filters
Channel Attention	Reduction ratio: 16
Spatial Attention	Kernel size: (7, 7), activation: Sigmoid
Global Average Pooling	–
Dense	128 units, activation: ReLU
Dropout	Rate: 0.5
Output (Dense)	14 units, activation: Softmax

As with previous models, the input time-series data were segmented into overlapping sliding windows of size $W = 32$, each containing 10 features as described in the preprocessing stage. Each window $X_i \in \mathbb{R}^{32\times10}$ was treated as a two-dimensional input, allowing the model to capture spatiotemporal dependencies ef-

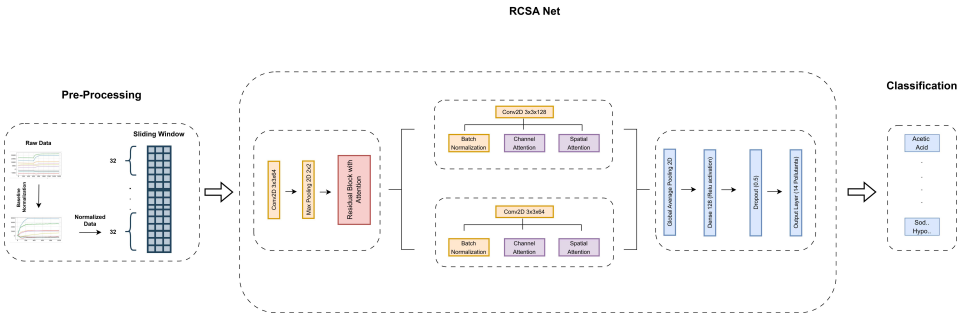


Figure 4.15: Proposed Residual Channel and Spatial Attention Network (RCSANet)

fectively.

$$X_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,F} \\ x_{i+1,1} & x_{i+1,2} & \cdots & x_{i+1,F} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i+W-1,1} & x_{i+W-1,2} & \cdots & x_{i+W-1,F} \end{bmatrix} \tag{4.11}$$

To improve learning capacity and reduce training instability in deeper networks, RCSANet employs two stacked residual blocks. Each block integrates:

- **Channel Attention:** A squeeze-and-excitation mechanism that assigns adaptive importance weights to each feature dimension using a reduction ratio of 16.
- **Spatial Attention:** A convolutional layer with a 7×7 kernel that emphasizes critical temporal regions within each windowed segment.

The architecture begins with a standard convolutional layer followed by batch normalization and max pooling. After the residual attention blocks, a global average pooling layer reduces the spatial dimensions, followed by a dense layer (128 units, ReLU activation), a dropout layer (rate = 0.5), and a final dense output layer with softmax activation to classify among the 14 pollutants. Through the integration of residual learning and dual-attention mechanisms, RCSANet was able to achieve improved discriminative capacity and training stability. Its performance surpasses that of baseline models by dynamically focusing on both spatial regions and feature dimensions critical to pollutant identification.

Chapter 5

Results and Discussion

This chapter presents the experimental evaluation and performance results of the smart sensing systems introduced in chapter 3. The focus is on assessing the classification capabilities of the SENSIPATCH wearable sensor for color sensing and the Smart Cable Water (SCW) system for water pollutant detection. The chapter is organized into two main sections. Section 5.1 presents the results of color classification using the SENSIPATCH system and PANTONE color samples. Section 5.2 reports the classification outcomes of water pollutants using the SCW system. In each case, performance metrics such as classification accuracy, confusion matrices, and cross-validation statistics are analyzed and discussed.

5.1 Color Sensing: Experimental Results and Analysis

This section presents the experimental results and performance analysis of the SENSIPATCH wearable sensor system for color classification using standardized PANTONE color samples. The goal of these experiments was to assess the system's ability to distinguish between various color classes based on spectral measurements acquired from physical PANTONE cards. Two experimental phases were conducted. The first was a preliminary evaluation involving a limited subset of PANTONE colors, aimed at exploring the feasibility of color classification using the SENSIPATCH system and informing modifications to the experimental design. The second phase was an extended study involving 100 PANTONE color samples, conducted under refined and standardized conditions. While the data preprocessing and traditional machine learning models remained consistent across both phases, the extended experiment incorporated additional deep learning archi-

tectures, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, to investigate their effectiveness in modeling spectral patterns for color classification. The results of both phases are presented in the following subsections, along with an analysis of classification performance using metrics such as accuracy, confusion matrices, and cross-validation results.

5.1.1 Preliminary Evaluation Using a Subset of PANTONE Colors

This subsection presents the classification results of machine learning models applied to spectral data acquired under two distinct sensor conditions: Exposed Sensor Condition (ESC) and Enclosed Sensor Condition (ENC). A 5-fold cross-validation approach, as shown in Figure 5.1, was adopted to assess model robustness and generalization. The performance metrics considered include classification accuracy, precision, and recall, offering a comprehensive view of the predictive effectiveness under varying environmental settings.

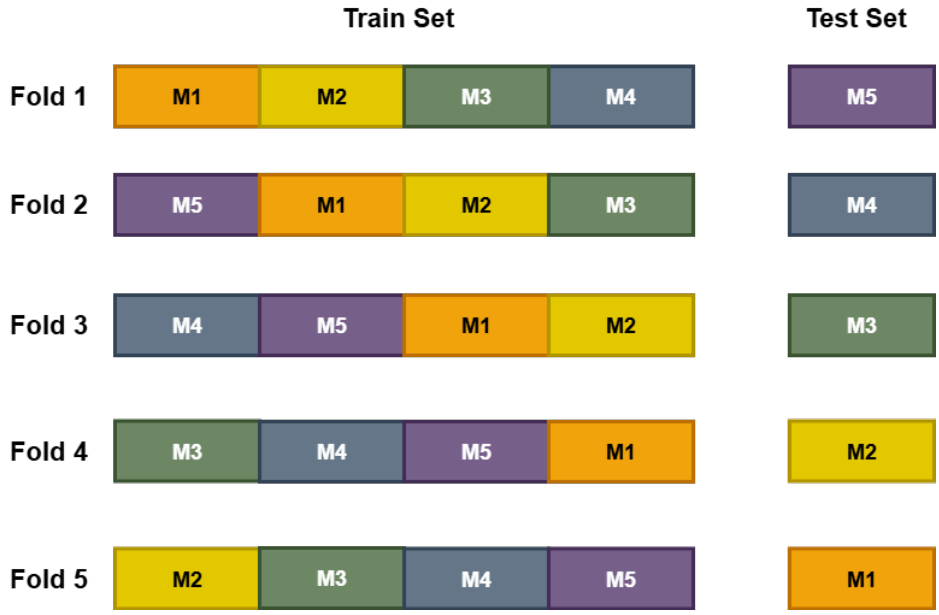


Figure 5.1: Stratified 5-fold cross-validation approach adopted. Each row represents the fold, and each column represents the measurement. In each fold, one measurement per class is used for testing, and the remaining four are used for training

Classification Performance under Exposed Sensor Condition (ESC)

In the ESC setup, the sensors were directly exposed to ambient environmental light, introducing potential variability and noise into the spectral readings. Table 5.1 summarizes the classification results for K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) models under this condition.

Among the tested models, MLP achieved the highest mean performance, with an average accuracy of 88.67%, precision of 82.32%, and recall of 87.03%. These results indicate that MLP was more resilient to environmental noise and able to extract relevant spectral features effectively.

In contrast, KNN and RF exhibited limited classification performance, with mean precision scores of 48.40% and 42.80%, respectively, suggesting susceptibility to false positives. Similarly, SVM recorded relatively low accuracy (44.20%) and precision (39.20%), reflecting its reduced discriminative power in noisy environments. The variability in model performance is likely due to interference from ambient lighting, which can affect sensor consistency. The MLP's robustness in this condition highlights its capability to learn discriminative representations even in the presence of environmental disturbances.

Classification Performance under Enclosed Sensor Condition (ENC)

In contrast to the ESC setup, the ENC condition involved enclosing the sensors within a controlled black-box environment, significantly minimizing exposure to ambient light. As shown in Table 5.2, all models benefited from the reduced environmental interference, with substantial improvements across all performance metrics.

Notably, the MLP model demonstrated superior consistency, achieving a mean accuracy of 91.63%, precision of 94.94%, and recall of 91.63%. It achieved near-perfect classification in some folds (up to 99.85% accuracy), emphasizing its capacity to exploit high-quality input data effectively. The SVM model also exhibited marked improvement (accuracy increased to 79.8%), highlighting its sensitivity to cleaner datasets.

These improvements validate the importance of environmental control in enhancing model performance and suggest that high-fidelity spectral data obtained through effective sensor shielding plays a crucial role in optimizing machine learning-based classification systems. The MLP confusion matrix under ENC (Figure 5.2) further illustrates strong class separation, although minor misclassifications remain.

Table 5.1: Classification performance for ESC

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean
KNN (Acc)	50%	60%	52%	56%	59%	55.40%
KNN (Prec)	39%	51%	52%	52%	48%	48.40%
KNN (Rec)	50%	60%	52%	50%	59%	54.20%
RF (Acc)	44%	54%	52%	43%	56%	49.80%
RF (Prec)	32%	43%	48%	42%	49%	42.80%
RF (Rec)	44%	54%	52%	38%	56%	48.80%
SVM (Acc)	40%	37%	50%	41%	53%	44.20%
SVM (Prec)	28%	31%	33%	55%	49%	39.20%
SVM (Rec)	40%	37%	50%	36%	53%	43.20%
MLP (Acc)	87.42%	99%	87.50%	68.40%	99%	88.67%
MLP (Prec)	81.15%	99%	81.25%	51.21%	99%	82.32%
MLP (Rec)	87.42%	99%	87.25%	62.50%	99%	87.03%

Table 5.2: Classification performance for ENC

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean
KNN (Acc)	37%	50%	35%	70%	81%	54.6%
KNN (Prec)	36%	48%	34%	68%	80%	53.2%
KNN (Rec)	37%	49%	35%	69%	81%	54.2%
RF (Acc)	49.22%	49.85%	64.16%	66.92%	79.28%	61.88%
RF (Prec)	48.90%	49.50%	63.80%	66.50%	78.90%	61.52%
RF (Rec)	49.10%	49.75%	64.00%	66.80%	79.10%	61.75%
SVM (Acc)	60%	75%	87%	95%	82%	79.8%
SVM (Prec)	59%	74%	86%	94%	81%	78.8%
SVM (Rec)	60%	75%	87%	95%	82%	79.8%
MLP (Acc)	74.72%	97.37%	99.7%	99.85%	86.52%	91.63%
MLP (Prec)	84.02%	97.75%	99.7%	99.85%	93.38%	94.94%
MLP (Rec)	74.72%	97.37%	99.7%	99.85%	86.52%	91.63%

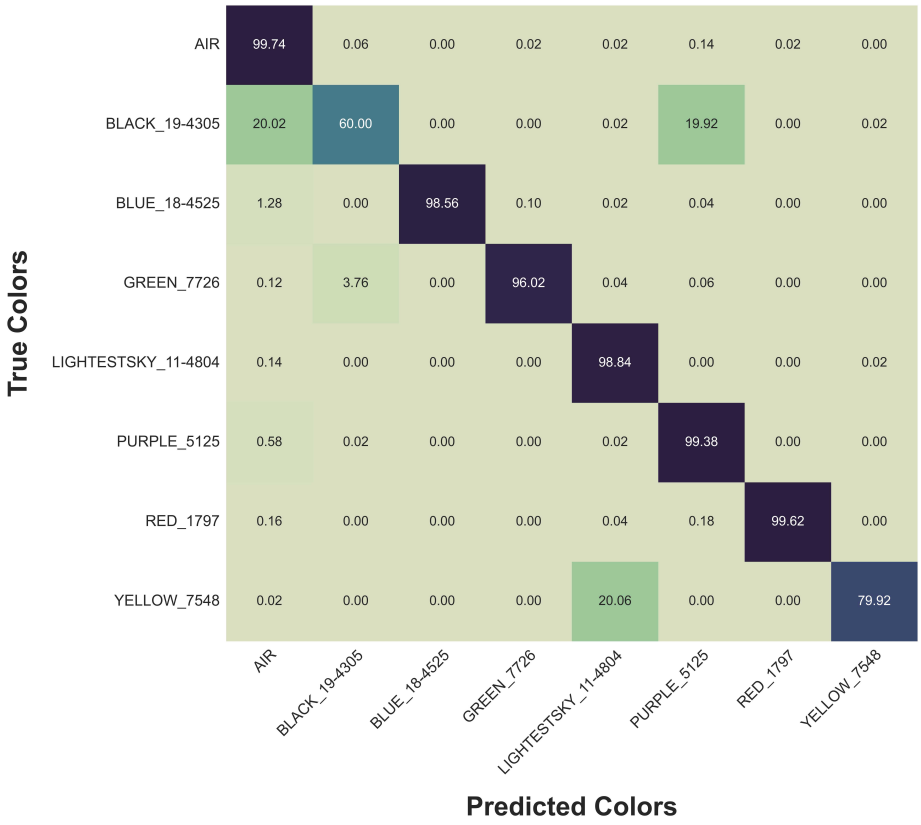


Figure 5.2: Confusion Matrix for MLP model under ENC condition

To further evaluate the consistency of the ESC classification models across folds, the mean accuracies and their corresponding standard deviations are summarized in Table 5.3. These values provide a measure of fold-to-fold stability, complementing the per-fold results previously reported.

Similarly, for the ENC dataset, the aggregated results are presented in Table 5.4, reporting both mean accuracies and standard deviations. This presentation style follows the reporting used in the Water Pollutant Detection experiments, ensuring consistency across sections.

Table 5.3: Mean and Standard Deviation of Classification Accuracies under ESC

Model	Mean Accuracy	Standard Deviation
KNN	55.40%	4.34%
RF	49.80%	5.93%
SVM	44.20%	6.91%
MLP	88.26%	12.51%

Table 5.4: Mean and Standard Deviation of Classification Accuracies under ENC

Model	Mean Accuracy	Standard Deviation
KNN	54.60%	20.31%
RF	61.89%	12.63%
SVM	79.80%	13.26%
MLP	91.63%	10.93%

5.1.2 Comprehensive Evaluation Using 100 PANTONE Colors

To evaluate the robustness and scalability of the SENSIPATCH system for spectral color classification, extensive experiments were conducted using 100 distinct PANTONE color samples under three experimental conditions: Firm Contact Condition With Weight (FCCW), Loose Contact Condition Without Weight (LCCW), and a Combined dataset integrating both scenarios. These conditions were designed to emulate both ideal and realistic sensor deployment situations and to assess model stability in the presence of variations in contact pressure. To ensure a robust and unbiased evaluation of model performance, 5-fold cross-validation was employed for both the FCCW and LCCW scenarios, as illustrated in Figure 5.3, while 10-fold cross-validation was applied to the Combined condition. These fold configurations were aligned with the number of available measurements per class: five measurements for both the FCCW and LCCW conditions, and ten measurements for the Combined condition (i.e., five from each contact type). In each fold, one measurement per class was held out for testing, while the remaining measurements were used for training. This approach preserved measurement-level independence and prevented data leakage, thus reducing the risk of overfitting. Additionally, this rotation-based fold strategy ensured that each model was evaluated on all measurement sequences.

For each of the 102 classes—comprising 100 PANTONE color cards, one white paper reference, and one baseline—five independent measurements were recorded under each contact condition. Each measurement initially consisted of 121 sen-

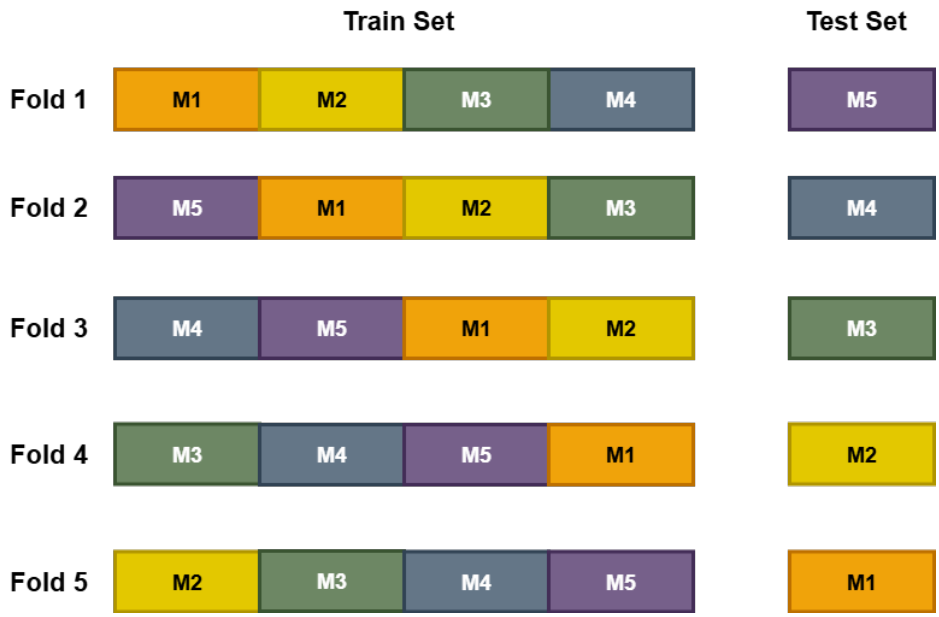


Figure 5.3: Stratified 5-fold cross-validation across 102 classes for LCCW and FCCW data. Each row corresponds to a fold, and each column represents a measurement. In each fold, one measurement per class is used for testing, and the remaining four are used for training.

sor readings, of which 100 were retained after baseline correction. To augment the dataset and enhance model robustness, bootstrapping was applied to each corrected measurement, resulting in 2,000 samples per measurement. This procedure produced a total of 1,020,000 samples per condition (102 classes × 5 measurements × 2,000 samples). In both FCCW and LCCW experiments, four out of the five measurements per class were used for training, while the remaining one was used for testing. Consequently, each condition included 8,000 training samples per class (4 measurements × 2,000 samples) and 2,000 test samples per class (1 measurement × 2,000 samples), resulting in a total of 816,000 training samples and 204,000 test samples per condition.

Table 5.5: Bootstrapped dataset structure, sample distribution, and cross-validation setup

Attribute	**FCCW**	**LCCW**	**Combined**
heightMeasurements per Class	5	10 (5+5)	2000
Samples per Measurement (after bootstrapping)	2000	816,000 (102 × 4 × 2000)	816,000 (102 × 4 × 2000)
Training Samples	2000	816,000 (102 × 4 × 2000)	816,000 (102 × 4 × 2000)
Test Samples	204,000 (102 × 1 × 2000)	204,000 (102 × 1 × 2000)	408,000 (102 × 2 × 2000)
Cross-Validation Folds	5	10	height

In the Combined condition, all ten measurements per class (five from FCCW and five from LCCW) were utilized. The training set comprised eight measurements per class—four from each contact condition—while the test set included the remaining two measurements (one from each condition). This configuration resulted in 1,632,000 training samples ($102 \times 8 \times 2{,}000$) and 408,000 test samples ($102 \times 2 \times 2{,}000$). By incorporating both contact configurations during both the training and evaluation phases, this setup allowed for a more comprehensive and realistic assessment of the model's generalization ability under varying physical deployment scenarios. Table 5.5 provides a summary of the measurement structure, sample distribution, and the corresponding cross-validation setup for each experimental condition.

Performance Under Firm Contact Condition With Weight (FCCW)

This scenario represents the most stable sensor configuration, where consistent pressure was applied via a custom ring-shaped weight to ensure direct and uniform contact between the sensor and the color card. This setup minimizes the impact of air gaps and supports high signal fidelity. As shown in Table 5.6, the SVM classifier demonstrated the best performance, achieving a mean accuracy of 97.29%, closely followed by RF and MLP, which attained 97.17% and 97.04%, respectively. While CNNs performed well with a mean accuracy of 96.17%, the LSTM model lagged significantly, with only 77.16% accuracy, indicating weaker performance in static conditions.

Table 5.6: Classification Accuracy Under FCCW (%)

Model	**Fold-1**	**Fold-2**	**Fold-3**	**Fold-4**	**Fold-5**	**Mean**
RF	98.09	96.99	98.70	96.34	95.75	**97.17**
SVM	96.58	97.62	98.64	96.12	97.49	**97.29**
MLP	96.85	96.53	98.61	96.17	97.05	**97.04**
LSTM	76.71	79.38	79.36	77.43	72.29	**77.16**
CNN	92.94	98.90	98.63	98.20	92.19	**96.17**

Table 5.7 presents additional evaluation metrics, precision, recall, and F1 score highlighting the consistency of SVM, MLP, and RF under optimal contact conditions. The LSTM model consistently underperformed across all evaluation metrics.

For the FCCW condition, the mean accuracies along with their standard deviations are reported in Table 5.8. These results provide insight into the stability of the classification models under full-contrast color settings.

Table 5.7: Mean Performance Metrics for ML Models under FCCW

Model	Precision	Recall	F1 Score
RF	97.53	97.17	96.91
MLP	97.31	97.04	96.74
SVM	97.48	97.29	97.09
CNN	94.83	96.17	95.11
LSTM	68.87	77.16	71.21

Table 5.8: Mean Accuracy and Standard Deviation under FCCW

Model	Mean Accuracy	Standard Deviation
CNN	96.17%	3.31%
LSTM	77.16%	2.90%
MLP	97.04%	0.94%
RF	97.17%	1.22%
SVM	97.29%	0.98%

Performance Under Loose Contact Condition Without Weight (LCCW)

This configuration simulates practical wearable usage scenarios where sensor-card alignment may be inconsistent due to motion or slight detachment. Without external pressure, the setup introduces variability in signal quality. Despite the increased variability, most models maintained strong performance as shown in Table 5.9. SVM led with a mean accuracy of 97.14%, followed by MLP 95.86%, and CNN 93.26%. RF showed slightly reduced performance 91.84%, while LSTM again performed the weakest 76.33%. As shown in Table 5.10, SVM once again demonstrated superior overall performance across all metrics. MLP remained highly reliable under varying contact conditions, underscoring its generalization capability.

Table 5.9: Classification Accuracy Under LCCW (%)

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean
RF	89.19	96.93	94.77	91.89	86.43	**91.84**
SVM	96.05	98.80	99.00	97.80	94.05	**97.14**
MLP	94.04	98.90	97.99	95.92	92.49	**95.86**
LSTM	74.73	79.29	78.60	77.77	71.27	**76.33**
CNN	91.77	92.01	94.61	92.81	95.13	**93.26**

The performance under LCCW is summarized in Table 5.11, where mean accuracies and standard deviations are presented to highlight both the central tendency

Table 5.10: Mean Performance Metrics for ML Models under LCCW

Model	Precision	Recall	F1 Score
RF	90.48	91.80	90.10
MLP	95.58	95.84	95.05
SVM	97.41	97.13	96.83
CNN	93.12	94.83	93.39
LSTM	69.39	76.33	70.97

and the variability across folds.

Table 5.11: Mean Accuracy and Standard Deviation under LCCW

Model	Mean Accuracy	Standard Deviation
CNN	93.26%	1.53%
LSTM	76.33%	3.32%
MLP	95.86%	2.67%
RF	91.84%	4.21%
SVM	97.14%	2.08%

Performance Under Combined Conditions (FCCW + LCCW)

To simulate a practical usage scenario with mixed contact stability, data from both FCCW and LCCW experiments were merged. This comprehensive dataset enables evaluation of model robustness across varying conditions. Classification was performed using 10-fold cross-validation as depicted in Figure 5.4. As summarized in Table 5.12, MLP achieved the best mean accuracy of 96.08%, followed by RF 94.96%. While SVM's accuracy dropped to 88.40%, CNN and LSTM showed reduced effectiveness with accuracies of 84.18%, and 70.21%, respectively.

Performance metrics shown in Table 5.13 confirm MLP's superior adaptability, with the highest F1 score of 95.37%, while SVM struggled with reduced recall under mixed-contact conditions.

Finally, the aggregated classification performance across combined conditions is presented in Table 5.14, showing the mean accuracies and associated standard deviations. This ensures a consistent comparison framework across all experimental settings.

5.1.3 Discussion of Color Sensing Results

The evaluation of the SENSIPATCH system for color classification began with a preliminary assessment using a small subset of PANTONE colors, carefully se-



Figure 5.4: Stratified 10-fold cross-validation across 102 classes adopted for combined data. Each row represents the fold, and each column represents the measurement. In each fold, two measurements per class are used for testing from FCCW and LCCW, and the remaining eight are used for training

lected. This phase was crucial in confirming the system's technical feasibility and in understanding baseline model behavior under controlled experimental conditions. The selected color samples were spectrally distinct, and measurements were acquired under two conditions i-e ESC and ENC. Results from this initial phase were encouraging. Machine learning models, particularly the MLP, achieved high classification accuracy, with most colors being identified correctly more than 95% of the time. Misclassifications were minimal and primarily limited to colors with closely overlapping reflectance characteristics. These outcomes laid a strong foundation for expanding the evaluation to more realistic and challenging scenarios. Figure 5.5 illustrates the mean classification accuracies of the machine learning

Table 5.12: Classification Accuracy for Combined Conditions (%)

Fold	RF	SVM	MLP	LSTM	CNN
1	98.90	85.93	97.96	65.51	85.27
2	97.72	87.19	97.63	64.38	79.43
3	97.82	87.05	97.74	62.66	84.57
4	93.19	85.65	95.52	67.59	80.61
5	93.19	84.11	94.29	71.81	77.86
6	88.87	85.86	89.27	71.05	84.24
7	98.43	92.56	97.26	76.21	85.91
8	97.40	93.72	99.30	74.74	88.69
9	93.34	93.75	97.46	74.38	91.33
10	89.47	88.14	88.14	73.84	83.89
Mean	**94.96**	**88.40**	**96.08**	**70.21**	**84.18**

Table 5.13: Mean Performance Metrics for ML Models for Combined Conditions (%)

Model	Precision	Recall	F1 Score
Random Forest	94.44	94.96	93.85
SVM	86.46	88.40	85.74
MLP	96.04	96.08	95.37
CNN	82.65	84.18	80.77
LSTM	61.77	70.21	62.81

models under both ESC and ENC conditions. As depicted, all models benefited from environmental enclosure, with the SVM showing the most substantial performance improvement. The MLP consistently achieved the highest accuracy in both scenarios, reaffirming its robustness to ambient light interference. As observed in Table 5.3 and Table 5.4, the MLP model consistently achieves the highest mean accuracy, albeit with relatively larger standard deviations, indicating some variability across folds.

The study was then extended to include all 100 PANTONE color classes, including plain white paper, encompassing a broader spectrum of colors with a wider variety of spectral similarities and differences. Importantly, data collection was conducted under two sensor contact conditions, i-e, Firm Contact Condition With Weight (FCCW) and Loose Contact Condition Without Weight (LCCW), to simulate a range of real-world deployment environments. By combining data from both settings into a unified dataset, the study aimed to improve the model's generalization across varied sensor placement and pressure conditions. Among all

Table 5.14: Mean Accuracy and Standard Deviation under Combined Conditions

Model	Mean Accuracy	Standard Deviation
CNN	84.18%	4.10%
LSTM	70.21%	4.83%
MLP	96.08%	3.82%
RF	94.96%	3.73%
SVM	88.40%	3.59%

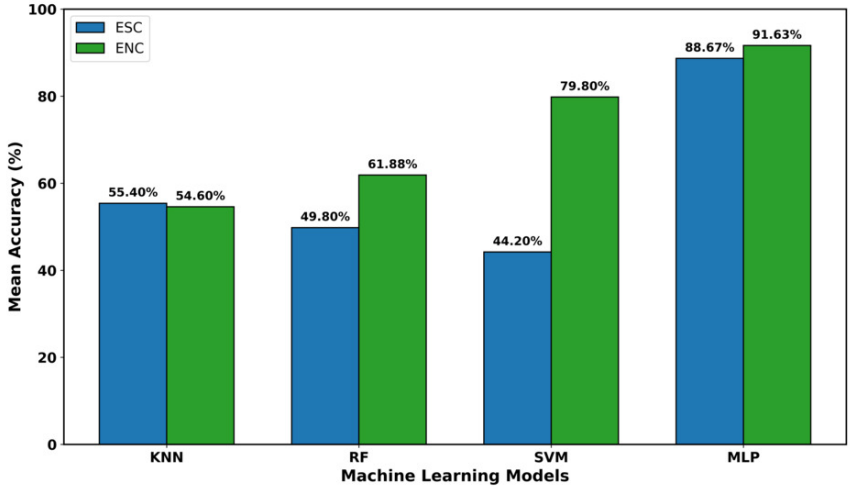


Figure 5.5: Comparison of mean classification accuracies of machine learning models under Exposed Sensor Condition (ESC) and Enclosed Sensor Condition (ENC).

models tested, MLP consistently delivered the most robust and accurate performance, achieving an average classification accuracy of 96.08% on the combined dataset. While the SVM slightly outperformed the MLP under the LCCW condition, and RF performed competitively under stable conditions, the MLP demonstrated greater overall balance between accuracy and robustness. CNN offered modest improvements in generalization, but LSTM models were less effective. As evident from Table 5.8, Table 5.11, and Table 5.14, the MLP model achieves the highest mean accuracies with relatively low variability, whereas the LSTM model shows lower accuracies and greater variability across folds.

Contact pressure emerged as a significant factor influencing classification stability. FCCW produced more clearly separated spectral patterns, resulting in improved classification across all models. That said, the performance degrada-

tion under LCCW was relatively minor, underscoring the sensor's and models' resilience in handling moderate real-world variability such as skin movement and misalignment. Notably, the combined dataset posed a greater challenge due to its heterogeneity, yet models trained on this dataset achieved improved generalization. This was particularly evident in fold-to-fold evaluation, where models trained on both FCCW and LCCW data handled unseen contact variations more effectively than those trained under a single condition. Given its balanced performance, deployment flexibility, and scalability, the MLP was selected as the primary classification model. Its compatibility with embedded optimization strategies, such as quantization and pruning, makes it well-suited for real-time wearable applications. Figure 5.6 presents the performance of the machine learning models under FCCW, LCCW, and combined contact conditions. The MLP maintained consistently high performance across all scenarios, while the SVM displayed a reduction in accuracy under the combined dataset. The LSTM model showed the lowest accuracy, further confirming its limited capability in handling spectral variability.

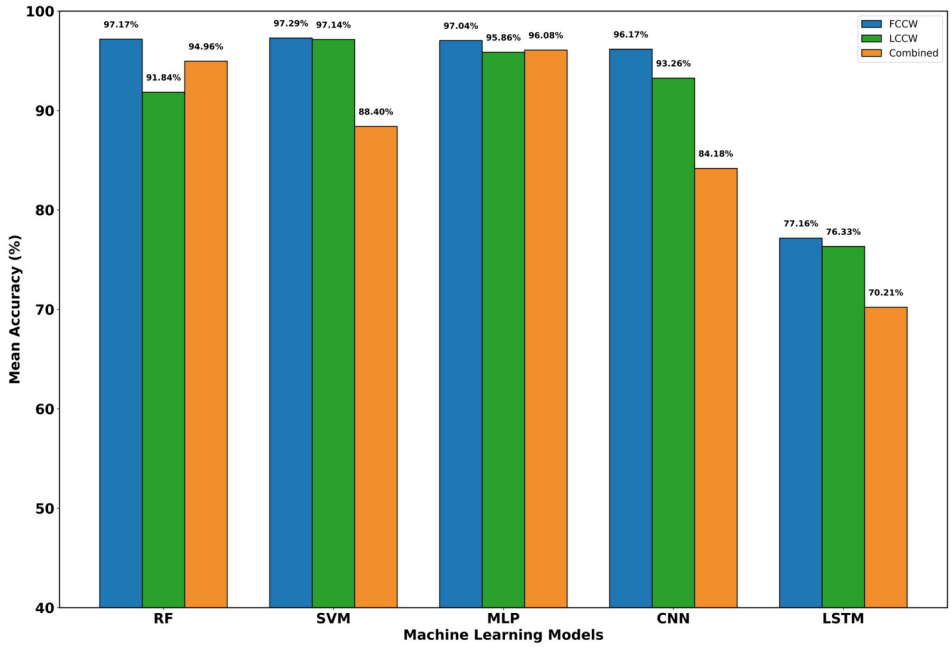


Figure 5.6: Classification performance of machine learning models under Firm Contact Condition With Weight (FCCW), Loose Contact Condition Without Weight (LCCW), and Combined Conditions.

Due to the space limitation, the full confusion matrix of 102 classes is not possible to add, so to explore per-class performance, two focused confusion matrices were generated one highlighting a subset of the top-performing color classes, as in Figure 5.7 and another depicting the subset of the lower-performing classes as in Figure 5.8.

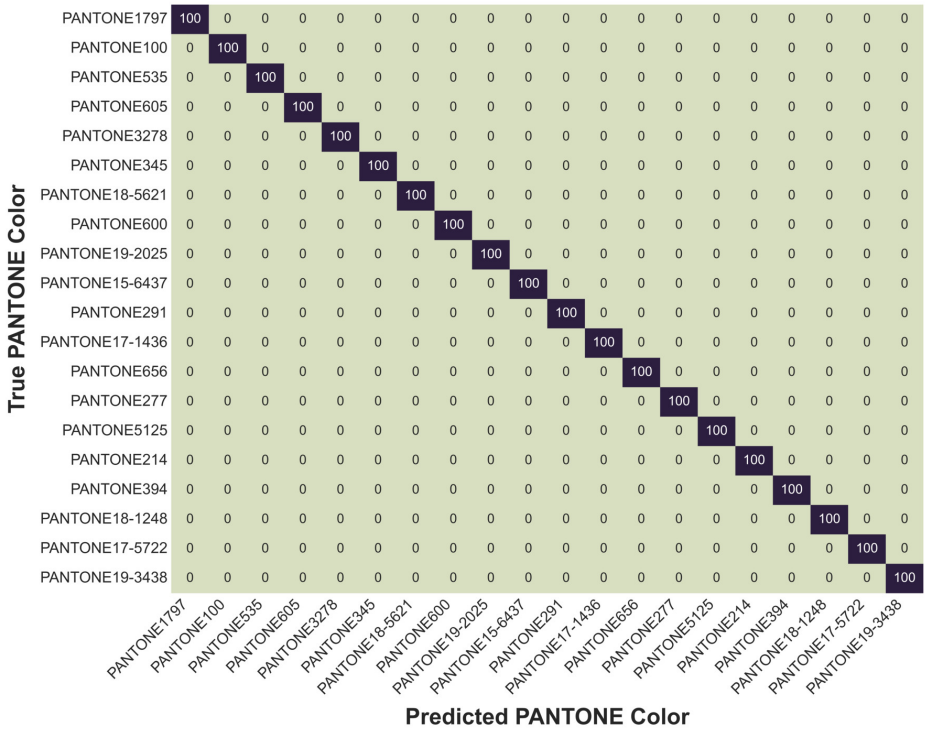


Figure 5.7: Subset confusion matrix of the MLP model trained on combined condition data, illustrating selected PANTONE color classes with high classification accuracies

The high-accuracy matrix included vivid, spectrally unique colors such as PANTONE1797, PANTONE345, and WHITE_PAPER, each classified with 100% accuracy. Conversely, the low-accuracy matrix identified a handful of more ambiguous tones like PANTONE166, PANTONE7720, and PANTONE17-1456, which were more susceptible to misclassification, especially when compared to spectrally similar neighbors e.g., PANTONE152, PANTONE18-4530. In some cases, dark tones were also misclassified as NO_PANTONE, likely due to their low reflectance signal

intensity.

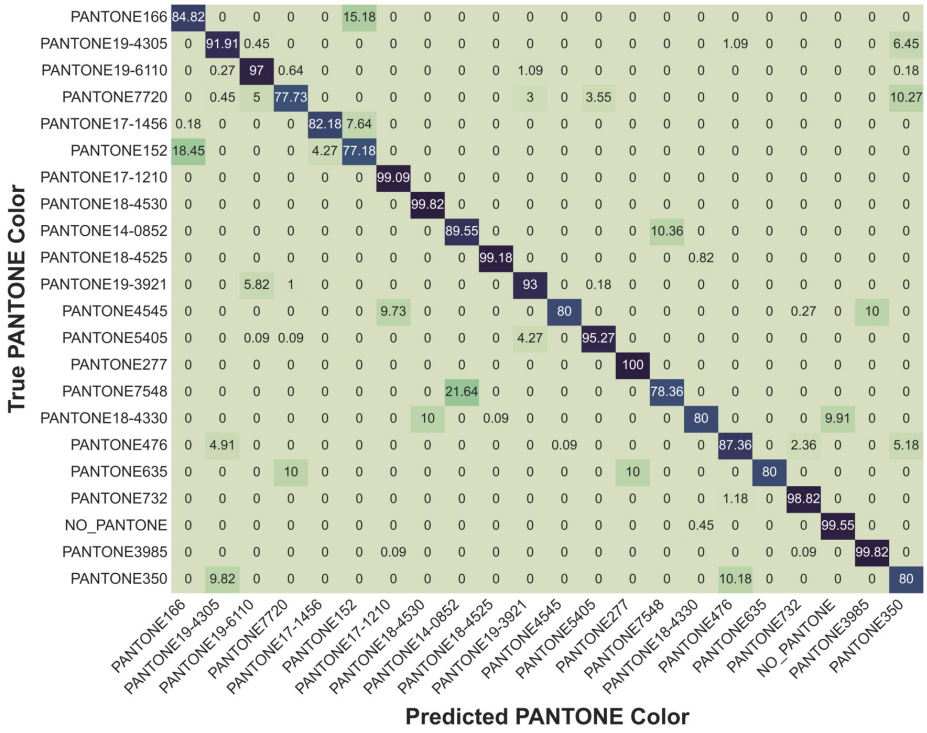


Figure 5.8: Subset confusion matrix for the MLP model trained on combined condition data, focusing on misclassifications among PANTONE color classes with the lowest accuracies

To support these observations, Table 5.15 provides a comprehensive breakdown of class-wise accuracy for all 102 classes. It confirms that 65 classes achieved an accuracy of 99% or higher, reflecting the model's strong ability to identify colors with unique spectral features. Particularly high accuracy was observed for saturated, spectrally distinct shades such as PANTONE100, PANTONE3278, and WHITE_PAPER. On the other hand, a small group of classes especially PANTONE7720, PANTONE166, PANTONE152, and PANTONE635, registered lower but still respectable accuracies in the 77–85% range, aligning well with the confusion matrix findings.

These results indicate a clear relationship between spectral separability and classification accuracy. When colors exhibit strong and distinct spectral signals, the

Table 5.15: MLP model Class-wise accuracy across Pantone color classes for combined conditions

Class	Accuracy (%)	Class	Accuracy (%)	Class	Accuracy (%)
NO_PANTONE	99.55	PANTONE100	100.0	PANTONE108	100.0
PANTONE11-4804	99.73	PANTONE12-0313	100.0	PANTONE12-0524	99.82
PANTONE12-0727	100.0	PANTONE12-5204	83.0	PANTONE13-0645	100.0
PANTONE13-0850	99.82	PANTONE13-1107	95.82	PANTONE13-5304	99.91
PANTONE14-0848	100.0	PANTONE14-0852	89.55	PANTONE14-1119	100.0
PANTONE14-1419	99.91	PANTONE14-1513	87.91	PANTONE14-1911	100.0
PANTONE14-3612	96.36	PANTONE14-4516	99.91	PANTONE142	99.82
PANTONE15-0543	100.0	PANTONE15-0942	90.0	PANTONE15-1247	100.0
PANTONE15-1816	99.09	PANTONE15-2706	99.91	PANTONE15-5210	89.91
PANTONE15-6437	100.0	PANTONE152	77.18	PANTONE16-1255	100.0
PANTONE16-1329	99.73	PANTONE16-1432	99.91	PANTONE16-1439	89.73
PANTONE166	84.82	PANTONE17-1210	99.09	PANTONE17-1436	100.0
PANTONE17-1456	82.18	PANTONE17-1564	97.36	PANTONE17-4021	99.82
PANTONE17-4524	99.91	PANTONE17-5722	100.0	PANTONE1797	100.0
PANTONE18-0119	86.18	PANTONE18-1248	100.0	PANTONE18-2326	99.91
PANTONE18-3025	97.91	PANTONE18-4330	80.0	PANTONE18-4525	99.18
PANTONE18-4530	99.82	PANTONE18-5621	100.0	PANTONE19-1250	98.36
PANTONE19-2025	100.0	PANTONE19-3438	100.0	PANTONE19-3921	93.0
PANTONE19-3935	98.36	PANTONE19-4305	91.91	PANTONE19-6110	97.0
PANTONE199	97.18	PANTONE214	100.0	PANTONE277	100.0
PANTONE291	100.0	PANTONE3278	100.0	PANTONE345	100.0
PANTONE350	80.0	PANTONE383	90.0	PANTONE394	100.0
PANTONE3985	99.82	PANTONE414	96.36	PANTONE4545	80.0
PANTONE466	89.91	PANTONE4685	91.36	PANTONE476	87.36
PANTONE485	98.64	PANTONE493	99.73	PANTONE5125	100.0
PANTONE535	100.0	PANTONE5405	95.27	PANTONE5555	96.73
PANTONE583	83.73	PANTONE600	100.0	PANTONE605	100.0
PANTONE635	80.0	PANTONE656	100.0	PANTONE677	99.82
PANTONE691	99.91	PANTONE732	98.82	PANTONE7415	99.18
PANTONE7493	100.0	PANTONE7548	78.36	PANTONE7590	100.0
PANTONE7654	91.36	PANTONE7720	77.73	PANTONE7726	100.0
PANTONE7737	100.0	PANTONE9041	100.0	PANTONE9044	99.0
PANTONE9180	99.91	PANTONE9241	90.36	PANTONE9320	99.82
PANTONECoolGray7	99.18	PANTONEWarmGray2	99.91	WHITE_PAPER	100.0

model is highly accurate. Where spectra are more similar or muted, classification becomes more challenging, but performance still remains robust. In summary, the progression from a small-scale, controlled evaluation to a comprehensive, real-world testing scenario confirms the scalability, reliability, and application readiness of the SENSIPATCH system. The selected MLP model, combined with the system's resilience to contact variability and spectral overlap, offers a strong foundation for practical deployment in wearable colorimetric sensing technologies.

5.2 Water Pollutant Detection: Experimental Results and Analysis

This section presents the evaluation results of the Smart Cable Water (SCW) system for water pollutant classification. The focus here is on assessing how effectively the system can distinguish between different chemical substances based on their

impedance characteristics, as outlined in the previous chapter. Two phases of experimentation were conducted to explore this: an initial evaluation involving a smaller set of pollutants, followed by a classification study with an expanded range of pollutants.

5.2.1 Experiment 1 Results

The performance of various deep learning models, including Transformers, RNN, LSTM, and CNN, was evaluated using ten-fold cross-validation. Table 5.18 summarizes the classification accuracy achieved by each model across the individual folds, along with their overall mean accuracies. Across the ten folds, CNN consistently delivered the best performance with a mean accuracy of 95.78%, followed by LSTM at 93.66%. The CNN model was particularly strong in Fold-4, where it reached a peak accuracy of 99.60%. LSTM also demonstrated strong and stable performance, outperforming the RNN, which recorded a lower mean accuracy of 85.28%. The reduced effectiveness of RNN is likely due to its vulnerability to the vanishing gradient issue, which hinders

Table 5.16: Classification Accuracy Across Ten Folds

Folds	**RNN**	**LSTM**	**CNN**	
Fold-1	75.43%	83.88%	85.58%	
Fold-2	79.53%	91.15%	94.21%	
Fold-3	92.42%	97.28%	97.48%	
Fold-4	88.04%	97.18%	99.60%	
Fold-5	89.47%	97.69%	98.26%	
Fold-6	83.90%	97.94%	97.94%	
Fold-7	94.33%	99.39%	99.23%	
Fold-8	85.21%	88.16%	95.37%	
Fold-9	82.91%	88.43%	92.89%	
Fold-10	81.64%	95.50%	97.26%	
Mean Accuracy	**58.68%**	**85.28%**	**93.66%**	**95.78%**

learning long-term dependencies.

Table 5.17: Mean and Standard Deviation of Classification Accuracies

Model	**Mean Accuracy**	**Standard Deviation**
RNN	85.28%	6.01%
LSTM	93.66%	4.73%
CNN	95.78%	2.27%

In addition to accuracy, model consistency across folds is an important consideration. Table 5.17 presents the standard deviation of the classification accuracies for each model. CNN had the lowest deviation (2.27%), indicating high stability across all folds. This can be attributed to CNN's ability to effectively extract local spatial patterns from the sliding window input. LSTM followed with a slightly higher deviation (4.73%), also suggesting robust performance. RNN exhibited more variability (6.01%), again reflecting its sensitivity to sequence length and gradient issues. The highest fluctuation was observed in Transformer models (14.38%), underscoring their inconsistency and relative unsuitability in this particular task context. Overall, CNN emerged as the most effective model for classifying the impedance signals collected in Phase 1, with LSTM also performing strongly. These findings support the suitability of CNNs for window-based signal classification tasks in water pollutant detection applications. As illustrated in the confusion matrix for this fold. Figure 6.6 shows the mean confusion matrix for 10 folds.

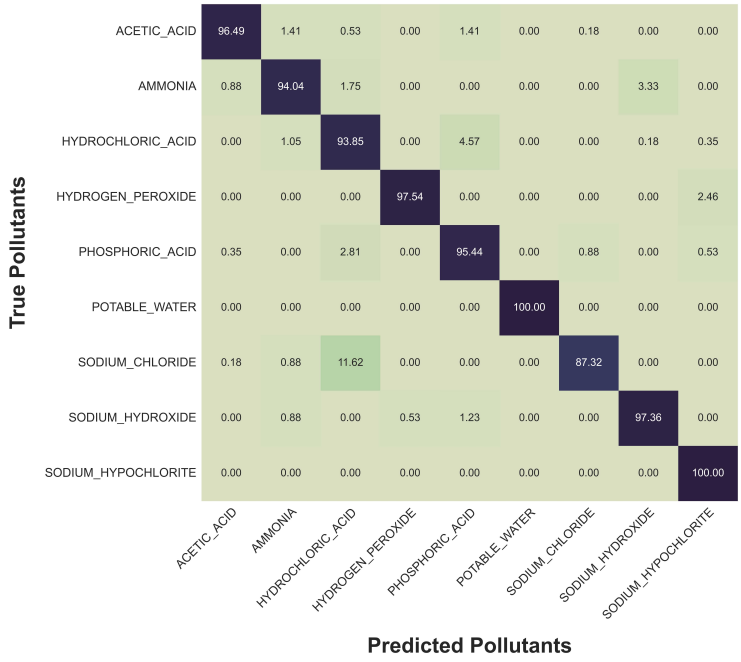


Figure 5.9: Mean Confusion matrix for CNN model

5.2.2 Experiment 2 Results

The classification performance of five deep learning models, RNN, LSTM, CNN, ResNet, and the proposed RCSANet was assessed through ten-fold cross-validation. The results are presented in Table 5.18, which details fold-wise accuracies along with overall mean performance. Among all evaluated models, RCSANet achieved the highest average accuracy of 94.21%, outperforming the baseline architectures. The improved performance of RCSANet is attributed to its use of both channel and spatial attention mechanisms, which enable the model to selectively focus on important features and spatial regions during training. Compared to ResNet, which attained a mean accuracy of 92.77%, the inclusion of these attention modules provided a measurable advantage. This architectural enhancement allowed RCSANet to better interpret and adapt to temporal variations in the impedance data collected during sequential pollutant injections. During data acquisition, pollutants were slowly injected into the water, creating brief transient fluctuations in the signal before stabilization occurred. Although normalization was applied to reduce baseline drift, some of these transient behaviors remained in the processed data. An example of this transient behavior is shown in Figure 4.11 for different pollutants. These brief signal variations posed a challenge for standard models such as RNNs and LSTMs, which are heavily dependent on temporal dependencies and often do not distinguish between transient noise and steady-state signal components [120, 121]. Similarly, although CNNs excel in extracting spatial patterns, they lack the temporal modeling capabilities needed for non-stationary time-series data [122]. ResNet offers hierarchical feature learning but cannot dynamically adjust to rapid changes in the data distribution [123]. In contrast, RCSANet demonstrated a clear advantage by dynamically attending to relevant signal regions. The channel attention component helped identify important features while suppressing irrelevant or noisy inputs introduced during transient phases. Simultaneously, spatial attention enabled the network to emphasize consistent spatial segments within the sliding windows, capturing the most discriminative signal patterns for accurate classification.

RCSANet also exhibited superior classification metrics beyond accuracy. Table5.19 outlines the precision, recall, and F1-score for each fold. The model demonstrated consistently high scores across all metrics, reflecting a balanced and reliable classification capability. Particularly strong performance was observed in Fold 5 and Fold 8, where near-perfect precision and recall were achieved. Slight performance dips in Fold 3 and Fold 6 are likely due to variations in sample distribution within those folds. Still, the model maintained overall robustness and generalizability across the dataset.

The mean confusion matrix for RCSANet, shown in Figure5.10, further confirms

Table 5.18: Classification Accuracy Across Ten Folds

Fold	RNN	LSTM	CNN	ResNet	RCSANet
Fold-1	68.64%	81.21%	89.06%	92.89%	93.01%
Fold-2	79.58%	83.15%	91.27%	91.06%	92.55%
Fold-3	85.44%	80.41%	74.97%	87.38%	86.74%
Fold-4	81.12%	91.57%	84.52%	98.78%	98.57%
Fold-5	86.26%	85.35%	97.26%	98.09%	99.73%
Fold-6	67.29%	80.72%	89.41%	88.40%	89.85%
Fold-7	68.29%	86.65%	93.47%	88.40%	95.41%
Fold-8	85.59%	93.70%	99.91%	93.27%	99.75%
Fold-9	75.98%	87.10%	96.27%	96.36%	93.36%
Fold-10	78.49%	92.98%	86.08%	92.62%	93.10%
Mean Accuracy	**77.67%**	**86.29%**	**90.22%**	**92.77%**	**94.21%**

its classification reliability across all pollutant categories. Misclassifications were rare and primarily involved chemically similar substances, such as *Acetic Acid* and *Formic Acid*, which exhibit close signal behavior.

In addition to accuracy and metric performance, the consistency of each model was evaluated using standard deviation (STD) of classification accuracies, presented in Table5.20. RCSANet demonstrated the highest stability, with a STD of only **2.13%**, indicating minimal variation across folds. This consistency highlights RCSANet's ability to generalize well across different training splits and maintain robustness to shifts in data distribution. By contrast, RNN and LSTM showed higher variability with STDs of 7.40% and 5.12%, respectively, suggesting less stable behavior.

In summary, the results from this phase highlight RCSANet as an accurate, stable, and generalizable model for water pollutant classification. Its attention-based architecture effectively handles transient signal noise and focuses on meaningful patterns, making it particularly well-suited for real-world sensing environments.

5.2.3 Discussion of Water Pollutant Detection Results

The experimental findings from both evaluation phases highlight the Smart Cable Water (SCW) system's ability to effectively classify chemical pollutants using impedance-based signals. While both experiments involved the use of deep learning models, there are important methodological distinctions between them particularly regarding data normalization and input structure that influenced performance outcomes. In the first experiment, transient signal regions occurring immediately after

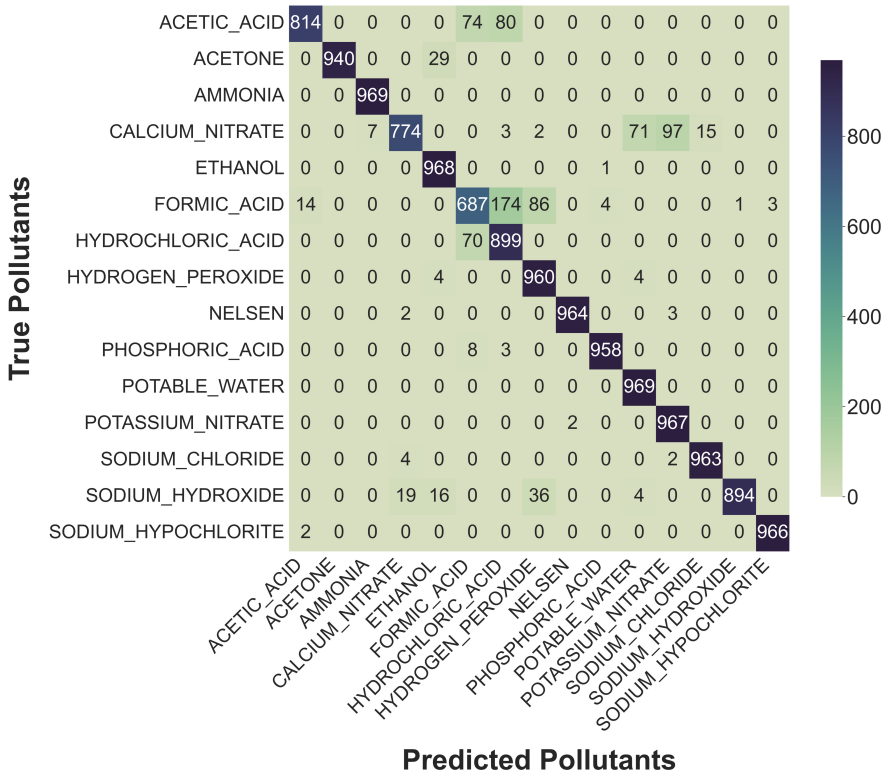


Figure 5.10: Mean Confusion Matrix for RCSANet

Table 5.19: Performance Metrics for RCSANet

Fold	Precision	Recall	F1-score
Fold-1	88.58%	93.01%	90.38%
Fold-2	89.26%	92.55%	90.33%
Fold-3	86.98%	86.74%	84.38%
Fold-4	98.72%	98.57%	98.56%
Fold-5	99.74%	99.73%	99.73%
Fold-6	92.90%	89.85%	87.83%
Fold-7	96.60%	95.41%	95.06%
Fold-8	99.76%	99.75%	99.75%
Fold-9	96.15%	93.36%	91.19%
Fold-10	96.18%	93.10%	90.85%
Mean	**94.55%**	**94.23%**	**92.80%**

Table 5.20: Mean Accuracy and Standard Deviation (STD) for Each Model

Model	Mean Accuracy	STD
RNN	77.67%	7.40%
LSTM	86.29%	5.12%
CNN	90.22%	4.05%
ResNet	92.77%	3.25%
RCSANet	94.21%	2.13%

each pollutant injection were deliberately removed during preprocessing. This resulted in cleaner, more stable signal segments of 600 data points being used for model training. The exclusion of transients allowed conventional models such as CNN and LSTM to focus on steady-state patterns, contributing to strong classification results, especially for CNN, which achieved a mean accuracy of 95.78%. In contrast, the second experiment retained the entire signal, including the transient regions, resulting in a longer input sequence of 1000 data points. This design decision better reflects real-world signal dynamics but introduces additional complexity due to the presence of unstable or noisy segments. As a result, models in this phase had to manage more variability and distinguish between meaningful trends and transient distortions. This added complexity was particularly well-handled by the proposed RCSANet model. Unlike conventional architectures, RCSANet leveraged both channel and spatial attention mechanisms to adaptively focus on stable, informative regions of the signal, while suppressing irrelevant or noisy fluctuations introduced during the transient phase. This architectural advantage enabled RC-

SANet to outperform all baseline models across accuracy, precision, recall, and F1-score metrics, and also maintain the lowest standard deviation across folds indicating consistent and robust generalization.

The comparison between experiments also underscores the limitations of models that lack attention mechanisms. RNNs and LSTMs, while able to capture sequential dependencies, were more sensitive to noisy transients in the second experiment, leading to higher variability in performance. Meanwhile, deeper convolutional architectures like ResNet performed better but still fell short of RCSANet due to their lack of dynamic focus on critical features. Overall, the results demonstrate that model choice must be closely aligned with the nature of the input data. When transient instability is present, as it often is in real-world applications, architectures like RCSANet that can adaptively filter and emphasize relevant signal patterns provide a significant performance edge. These findings not only validate the SCW system's design but also highlight the importance of model adaptability in handling non-stationary sensor data for reliable water quality monitoring.

Chapter 6

Battery State of Charge Estimation Using EIS and Deep Learning

This Chapter explores the application of deep learning techniques for the estimation of battery State of Charge (SoC) using Electrochemical Impedance Spectroscopy (EIS). In this chapter, a data-centric approach is presented, wherein EIS measurements collected from lithium iron phosphate (LFP) batteries are transformed into image-based representations using Gramian Angular Field (GAF) encoding. This transformation enables the application of convolutional neural networks (CNNs) to perform SoC classification based on frequency-domain impedance signals. The chapter details the experimental setup for data acquisition, the preprocessing steps applied to convert raw EIS signals into usable input formats, the design of deep learning architectures, and the evaluation metrics used to assess model performance.

6.1 Dataset Generation

To facilitate accurate and scalable data-driven analysis for battery state estimation, a novel dataset [32] was developed based on Electrochemical Impedance Spectroscopy (EIS) measurements on Lithium Iron Phosphate (LFP) batteries. The objective of this dataset generation process was to capture fine-grained impedance signatures across a wide range of State of Charge (SoC) levels using controlled and repeatable experimental procedures.

6.1.1 Battery Selection and Configuration

A total of eleven 600 mAh cylindrical LFP cells were employed in this study. The selected cells are of model IFR14500EC, manufactured by O'Cell New Energy Technology Co., Ltd. Each cell underwent two full discharging cycles with impedance measurements taken across a range of state-of-charge (SoC) values.

The key specifications of the batteries used in this study are detailed in Table 6.1.

Table 6.1: Characteristics of the adopted batteries.

	Battery Characteristics
Manufacturer	O'Cell New Energy Technology CO., LTD
Model	IFR14500EC
Type	Cylindrical
Nominal Capacity	600 mAh
Nominal Voltage	3.2 V
Charge Voltage	3.65 V
Charge Current	300 mA (0.5C)
Cut–off Current	30 mA (0.05C)
Cut–off Voltage	2.0 V

6.1.2 Data Acquisition System

The experimental platform was custom-designed to meet the specific requirements of this study. Its main features include:

- controlled discharge of a single battery at a time,
- SoC resolution finer than 0.1%, and
- automated impedance measurements at each defined SoC decrement (ΔSoC).

Instrumentation and Configuration

The system uses the *Hioki BT4560* impedance analyzer for EIS measurements. It injects an AC current into the battery and measures the voltage response to compute the impedance. Depending on the impedance range, it applies different root mean square current levels: 150 mA (up to 30 mΩ), 50 mA (up to 300 mΩ),

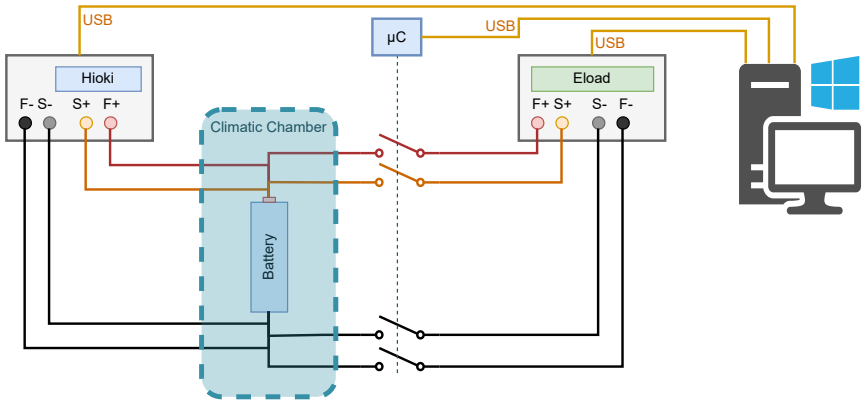


Figure 6.1: Schematic diagram of the implemented experimental setup.

and 5 mA (up to 3 Ω). A four-terminal measurement configuration is used to minimize the effect of contact and wire resistances.

Battery discharge is managed by the *ZKETECH EBD-A20H* programmable electronic load, which supports discharge currents up to 20 A. It also uses a four-wire connection, ensuring voltage accuracy is unaffected by cable resistances. A microcontroller-controlled switching circuit allows disconnection of the electronic load during EIS measurements, maintaining signal integrity. Although each switch introduces approximately 100 mΩ of contact resistance, the four-terminal setup compensates for this.

Error Minimization Strategies

To reduce external influences on measurements:

- the physical setup was fixed before all tests, ensuring consistent geometry and cable arrangement;
- impedance meter calibration was performed whenever a new battery was installed, compensating for offsets and environmental changes;
- four-terminal connections were used for both the load and analyzer, minimizing contact and lead resistance effects.

Measurement Uncertainty

Measurement accuracy was evaluated using the ISO ENV 13005 “Guide to the Expression of Uncertainty in Measurement (GUM)” [124]. Based on Type B eval-

uation:

- the *Hioki BT4560* had an uncertainty of less than 5 mΩ for both the real and imaginary impedance components,
- the *ZKETECH EBD-A20H* showed an uncertainty of 0.6 mAh in capacity measurements.

Environmental Control and Automation

All tests were conducted in a Peltier-based climatic chamber, which maintained a constant temperature of 20 °C with an accuracy of ±1 °C, as specified by the manufacturer. A built-in sensor monitored and regulated the temperature to avoid thermal drift in the measurements.

The entire setup was automated through a Python-based software platform that accepts user-defined parameters such as discharge current, cut-off voltage, number of impedance measurements, and stimulus frequencies. It manages SoC tracking, switching, data logging, and communication with all connected devices.

A total of eleven LFP batteries were tested using this system. Their main characteristics are summarized in Table 6.1.

6.1.3 Measurement Protocol

Before the EIS tests were carried out, the actual capacity of each battery was evaluated using the Coulomb Counting method. The procedure began with placing the battery inside a climate-controlled chamber set to the test temperature. After a rest period of 10 hours—sufficient to allow thermal stabilization—the battery was deemed ready for capacity evaluation. At this point, the battery was fully charged using a constant current–constant voltage (CC–CV) method as specified by the manufacturer (see Table 6.1). The battery was then discharged at a constant current of $0.5\,C$, matching the discharge current used in subsequent impedance measurements. The electronic load recorded the actual deliverable capacity of the battery, which may differ from the nominal value due to manufacturing tolerances, material inconsistencies, or calendar aging. This capacity evaluation was performed only once prior to the main test sequence, as the manufacturer guarantees a cycle life exceeding 2000 under more demanding test conditions than those applied here.

The complete measurement protocol is illustrated in Figure 6.2. It consists of two primary phases: battery charging (highlighted in magenta) and battery discharging (highlighted in blue).

The charging phase follows the CC–CV method: the battery is first charged at a constant current until the terminal voltage reaches the upper limit V_{charge}, after

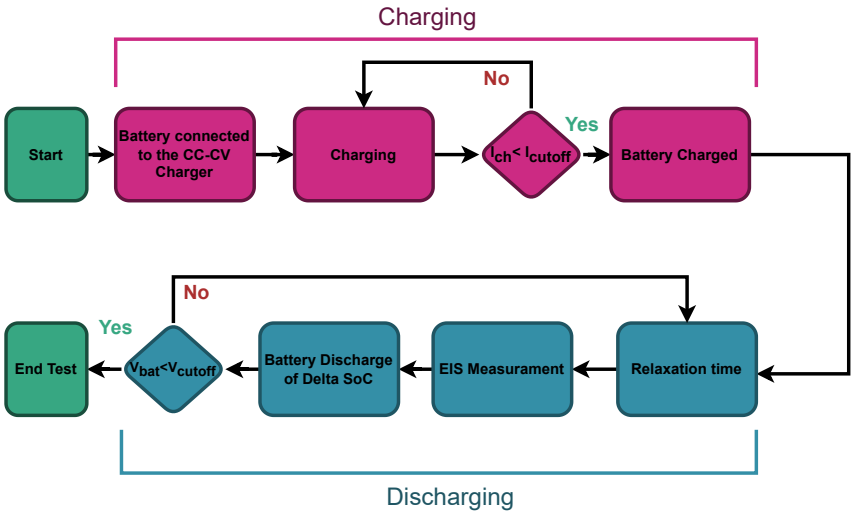


Figure 6.2: Adopted procedure to obtain the experimental dataset.

which the voltage is held constant while the current naturally decreases. Charging concludes when the current drops below a threshold value $I_{\text{cut-off}}$, indicating full charge. After the charging step, a relaxation period of one hour allows electrochemical stabilization. The battery is disconnected from the electronic load using a relay system, and impedance measurements are taken at all predefined frequencies. The resulting EIS data is stored on the control PC.

Once measurements are complete, the battery is discharged by a predefined energy increment corresponding to a specific SoC interval, using a discharge rate of $0.5\,C$. Following each discharge step, a relaxation time of 15 minutes is observed before the next impedance measurement is conducted. This cycle repeats until the battery voltage falls below the cut-off voltage $V_{\text{cut-off}}$, at which point the test terminates. The parameters V_{charge}, $I_{\text{cut-off}}$, and $V_{\text{cut-off}}$ are all specified by the battery manufacturer and detailed in Table 6.1. The state of charge (SoC) is defined as the ratio of the battery's remaining usable energy to its full deliverable energy under current test conditions, which depend on factors like temperature, discharge rate, and battery health.

To ensure accurate SoC labeling, a separate discharge cycle without EIS measurements was conducted for each battery at the test current and temperature. The actual battery capacity obtained from this cycle, via Coulomb Counting, was used to calculate the ΔSoC step, defined as:

$$\Delta\text{SoC} = \frac{\text{Measured Capacity}}{\text{Number of SoC Levels}}.$$

This protocol ensures a high-quality, granular EIS dataset that faithfully captures battery behavior across multiple SoC levels under controlled conditions.

6.1.4 Dataset Composition and File Format

The experimental dataset consists of two main types of data: Electrochemical Impedance Spectroscopy (EIS) measurements and battery capacity evaluations. Both data types are stored in standard comma-separated values (CSV) format to ensure compatibility with common data analysis tools.

EIS Measurement Files

For each test cycle, impedance spectra were recorded at 20 different SoC levels, ranging from 5% to 100% in 5% increments. Each CSV file contains impedance measurements across 28 frequency points. The structure of these files includes six columns, as detailed in Table 6.2.

Table 6.2: EIS measurement file structure and column descriptions.

Output Parameter	Column Name	Range	Unit	Description
Frequency	`Frequency`	0.01–1000	Hz	Frequency used for impedance measurement
Real Impedance	`R`	Context-dependent	Ω	Real part of the complex impedance
Imaginary Impedance	`X`	Context-dependent	Ω	Imaginary part of the complex impedance
Battery Voltage	`V`	0–20	V	Battery voltage prior to EIS measurement
Temperature	`T`	Based on chamber setting	°C	Ambient temperature during testing
Measurement Range	`Range`	0.3–3	Ω	Internal range setting of impedance analyzer

Capacity Measurement Files

Each battery was also characterized through a capacity evaluation test. The corresponding CSV files contain time-series data including voltage and current during the discharge process. Table 6.3 describes the format and meaning of these entries. These structured datasets enable systematic preprocessing, transformation, and modeling for downstream machine learning applications in SoC estimation.

6.1.5 Nyquist Plot Analysis of Impedance Data

Electrochemical Impedance Spectroscopy (EIS) data is often visualized using Nyquist plots, which represent the imaginary part of the impedance ($Im(\dot{Z})$) against the

Table 6.3: Capacity measurement file structure and column descriptions.

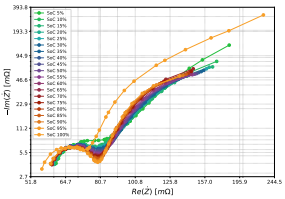
Column Name	Unit	Description
Time	s	Elapsed time during the discharge test
Cur	A	Discharge current
Vol	V	Terminal voltage of the battery

real part ($Re(\dot{Z})$). This graphical representation is useful for assessing the electrochemical behavior of the battery and identifying anomalies or outliers in the measurements. For each battery, a Nyquist plot was generated at every acquired SoC level. Figure 6.3 shows the impedance spectra from the first test cycle for all eleven batteries. Each subplot includes the 20 impedance curves corresponding to SoC levels ranging from 5% to 100%.

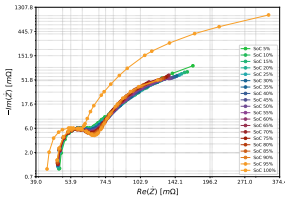
An observable trend across all batteries is that the impedance spectrum at 100% SoC diverges significantly from the others. This deviation is consistent with electrochemical theory. At high SoC levels, lithium-ion concentration in the cathode (positive electrode) is minimal, which can impede ion diffusion. As a result, the characteristic diffusion time becomes very large, leading to elevated impedance values [**huang2016impedance**]. Aside from this high-SoC behavior, the general impedance profiles align with what has been reported in the literature for LFP cells of similar chemistry, design, and capacity [**8915538**, **9072657**]. This visual validation confirms the consistency and electrochemical realism of the acquired dataset, making it suitable for downstream modeling and learning tasks.

6.1.6 Battery SoC Estimation Using GAF-Based Image Encoding

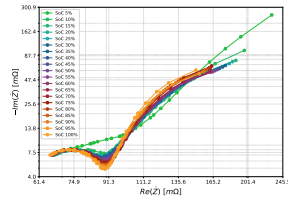
This section describes an image-based approach for estimating the State of Charge (SoC) of lithium-based batteries using Electrochemical Impedance Spectroscopy (EIS) data. Instead of processing the impedance values in their original numerical format, the measurements are converted into two-dimensional image representations using the Gramian Angular Field (GAF) [33] encoding. These encoded images are then used as input to deep learning models for SoC classification. The GAF method, commonly applied to time series data, is adapted here to represent frequency-domain impedance patterns as spatially structured images. This transformation enables the application of convolutional neural networks (CNNs), which are effective at extracting local and global spatial features, to classify SoC levels. An overview of the adopted methodology is shown in Figure 6.4. It summarizes the workflow from EIS data acquisition and GAF transformation to image-based



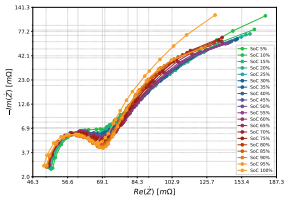
(a) First battery (B01)



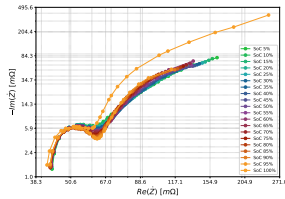
(b) Second battery (B02)



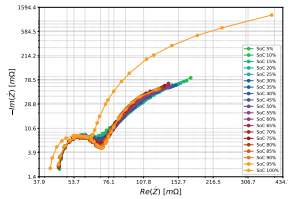
(c) Third battery (B03)



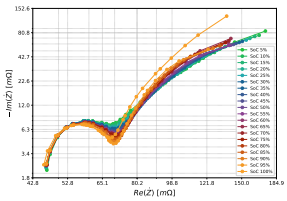
(d) Fourth battery (B04)



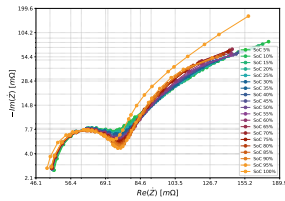
(e) Fifth battery (B05)



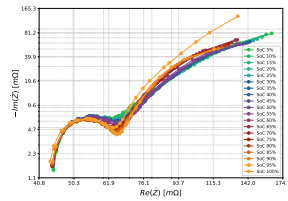
(f) Sixth battery (B06)



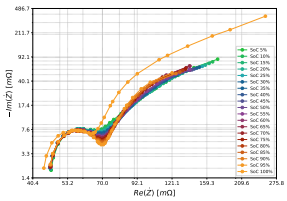
(g) Seventh battery (B07)



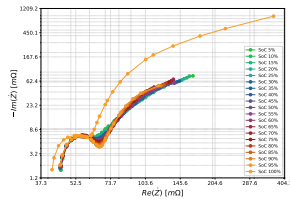
(h) Eighth battery (B08)



(i) Ninth battery (B09)



(j) Tenth battery (B10)



(k) Eleventh battery (B11)

Figure 6.3: Nyquist plots of the eleven batteries from the first test cycle. Each curve corresponds to one of the 20 acquired SoC levels.

SoC estimation using deep learning models. The subsequent sections explain the encoding process, dataset preparation, model architecture, and evaluation procedures.

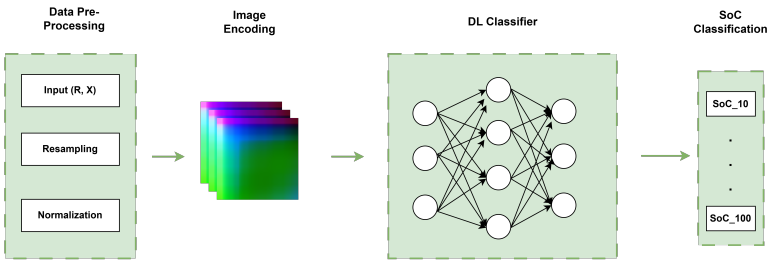


Figure 6.4: Overview of the adopted methodology for SoC estimation using GAF-encoded EIS data.

6.1.7 Transformation of EIS Data Using Gramian Angular Fields

To enable battery State of Charge (SoC) estimation via deep learning, Electrochemical Impedance Spectroscopy (EIS) data were converted into two-dimensional image representations using the Gramian Angular Field (GAF) method [33]. This encoding approach, originally developed for time-series data, is here adapted to impedance spectra—transforming frequency-ordered sequences of complex values into structured images suitable for input to convolutional neural networks (CNNs). Although EIS data is not temporal, the ordered nature of impedance values across frequencies permits the use of methods like GAF, which rely on structured sequences. Each sample consists of 28 impedance values (real and imaginary parts) measured across frequency. These were first linearly interpolated to produce sequences of length $N = 100$ to standardize input size for GAF computation. This step ensures uniform image dimensions and enables consistent feature encoding.

The interpolated data vectors were then normalized to the range $[-1, 1]$ to enable angular transformation:

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \tag{6.1}$$

Each normalized value $\tilde{x}_i \in [-1, 1]$ is mapped to a polar coordinate using the following transformation:

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), \\ r_i = \frac{i}{N}, \quad i = 1, \ldots, N \end{cases} \tag{6.2}$$

The angular component ϕ_i retains the normalized magnitude information, while the pseudo-radial coordinate r_i is proportional to the index of the data point. These polar coordinates are used to construct two Gramian matrices: one using the cosine of angular sums, the other using the sine of angular differences. The full forms of these matrices are given below.

$$G_{cos} = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \ldots & \cos(\phi_1 + \phi_N) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \ldots & \cos(\phi_2 + \phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_N + \phi_1) & \cos(\phi_N + \phi_2) & \ldots & \cos(\phi_N + \phi_N) \end{bmatrix} \tag{6.3}$$

$$G_{sin} = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \sin(\phi_1 - \phi_2) & \ldots & \sin(\phi_1 - \phi_N) \\ \sin(\phi_2 - \phi_1) & \sin(\phi_2 - \phi_2) & \ldots & \sin(\phi_2 - \phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\phi_N - \phi_1) & \sin(\phi_N - \phi_2) & \ldots & \sin(\phi_N - \phi_N) \end{bmatrix} \tag{6.4}$$

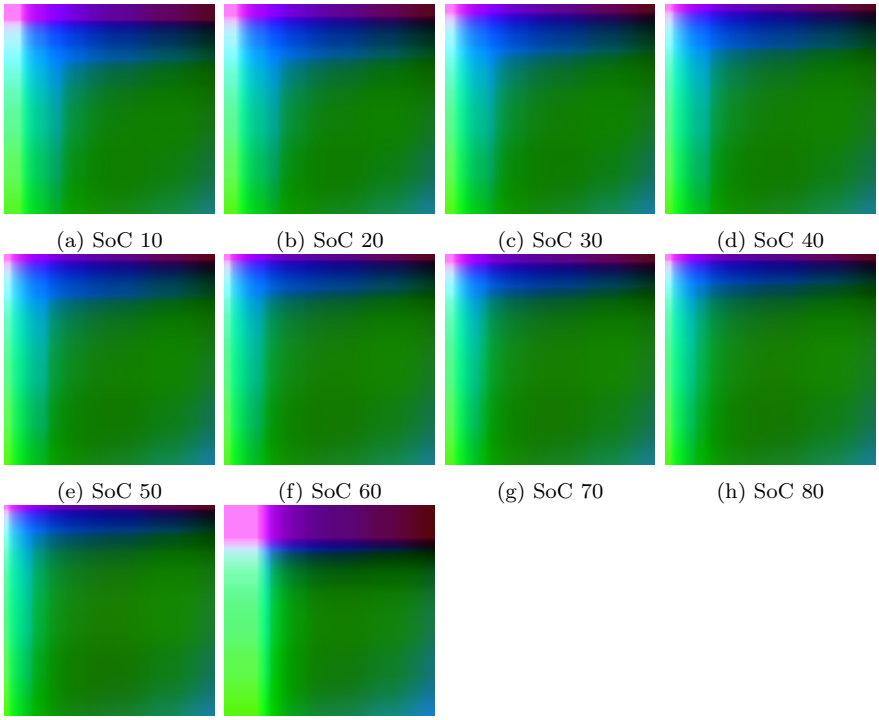
These matrices capture distinct geometric properties: G_{cos} encodes global phase interactions via angular summation, while G_{sin} highlights differential changes between frequency points. Both offer complementary views of impedance patterns. To convert EIS samples into RGB images suitable for CNNs, three Gramian matrices were selected to populate the image channels:

- **Red channel:** $G^X_{\cos}$ (cosine of imaginary part)
- **Green channel:** $G^X_{\sin}$ (sine of imaginary part)
- **Blue channel:** $G^R_{\cos}$ (cosine of real part)

A representative example of GAF-encoded images for battery B01 across SoC levels is shown in Figure 6.5.

This configuration ensures the resulting image reflects both magnitude and phase characteristics of the complex impedance data across frequencies. After generating GAF images for all 20 SoC levels, a label aggregation step was applied to simplify classification and balance the dataset. The original 20 levels were grouped into 10 broader SoC classes according to the following rule:

- SoC 5% → Class SoC5
- SoC 10% → Class SoC10



(a) SoC 10 (b) SoC 20 (c) SoC 30 (d) SoC 40

(e) SoC 50 (f) SoC 60 (g) SoC 70 (h) SoC 80

(i) SoC 90 (j) SoC 100

Figure 6.5: GAF-encoded images for battery B01 at different SoC levels (after label aggregation).

- SoC 15% and 20% → Class SoC20
- SoC 25% and 30% → Class SoC30
- SoC 35% and 40% → Class SoC40
- SoC 45% and 50% → Class SoC50
- SoC 55% and 60% → Class SoC60
- SoC 65% and 70% → Class SoC70
- SoC 75% and 80% → Class SoC80
- SoC 85% to 100% → Class SoC100

This relabeling reduces class imbalance, limits overfitting, and improves learning stability—especially critical given the modest dataset size. GAF images were generated for each of the 10 aggregated classes across 10 different batteries.

6.1.8 Deep Learning-Based Classification of GAF-Encoded EIS Data

To evaluate the effectiveness of image-based deep learning for SoC classification, this study employs two convolutional neural network (CNN) architectures: a custom-designed Simple CNN [125] and the well-established AlexNet model [126]. These models are tasked with classifying the State of Charge (SoC) of lithium iron phosphate (LFP) batteries based on GAF-encoded representations of EIS data. The primary objective is to determine the ability of these networks to learn discriminative features from the impedance-based image data. To ensure robust and unbiased performance evaluation, a 10-fold cross-validation scheme was applied. The dataset consisted of measurements from ten individual batteries (B01 through B10), and each fold of the cross-validation involved a unique split of the batteries into training, validation, and test sets.

Cross-Validation Protocol

The adopted cross-validation strategy ensures that each battery appears exactly once in the test set, twice in the validation set, and six times in the training set over the 10 folds. This systematic rotation helps mitigate overfitting, promotes generalization, and maximizes the use of all available data.

- **Training Set:** In each fold, data from six batteries were allocated for training the models. For instance, in the first fold, batteries B01 to B06 were used for training. Across all folds, each battery appeared in the training set six times.

- **Validation Set:** Two batteries were used for model validation in each fold. These validation sets served to tune hyperparameters and monitor model performance during training. Each battery was part of the validation set in exactly two folds.
- **Testing Set:** The remaining two batteries in each fold were reserved for final testing. This independent test set ensured that model performance was evaluated on unseen data. Each battery contributed once to the test set across the 10 folds.

This approach ensured comprehensive use of the available data while preserving independent evaluation and fairness in model comparison.

Network Architectures

The two evaluated architectures—the custom Simple CNN and AlexNet—were configured to extract features from GAF-encoded images for multi-class classification. The detailed layer configurations are presented in Tables 6.4 and 6.5, respectively. Both networks were optimized using the Adam optimizer, with learning rates of 0.001 for the Simple CNN and 0.0001 for AlexNet. The categorical cross-entropy loss function was used to guide training, consistent with the multi-class nature of the SoC classification task.

Table 6.4: Simple CNN architecture.

Layer Type	**Configuration**	**Activation Function**
Conv2D	32 filters, kernel size: 5x5	ReLU
Conv2D	64 filters, kernel size: 5x5	ReLU
Conv2D	128 filters, kernel size: 5x5	ReLU
Conv2D	256 filters, kernel size: 3x3	ReLU
MaxPooling2D	Pool size: 2x2, stride: 2	-
Conv2D	512 filters, kernel size: 3x3	ReLU
GlobalAvgPooling2D	-	-
Dense	128 units	ReLU
Dropout	Rate: 0.5	-
Dense	64 units	ReLU
Dense	10 units (output layer)	Softmax

Training and Validation Strategy

Model training followed the 10-fold cross-validation design. Each training phase involved early stopping based on validation performance to mitigate overfitting.

Table 6.5: AlexNet architecture.

Layer Type	Configuration	Activation Function
Conv2D	96 filters, kernel size: 11x11	ReLU
BatchNormalization	-	-
MaxPooling2D	Pool size: 2x2	-
Conv2D	256 filters, kernel size: 5x5	ReLU
BatchNormalization	-	-
MaxPooling2D	Pool size: 2x2	-
Conv2D	512 filters, kernel size: 3x3	ReLU
BatchNormalization	-	-
MaxPooling2D	Pool size: 2x2	-
Conv2D	1024 filters, kernel size: 3x3	ReLU
BatchNormalization	-	-
Conv2D	1024 filters, kernel size: 3x3	ReLU
BatchNormalization	-	-
MaxPooling2D	Pool size: 2x2	-
Flatten	-	-
Dense	1024 units	ReLU
Dropout	Rate: 0.5	-
Dense	4096 units	ReLU
Dropout	Rate: 0.5	-
Dense	10 units (output layer)	Softmax

The use of multiple validation batteries in each fold ensured reliable generalization estimates, while consistent testing on held-out batteries supported unbiased comparisons.

Model Performance Evaluation

Model evaluation focused primarily on classification accuracy and categorical cross-entropy loss across all folds. To gain a deeper understanding of model behavior across SoC levels, confusion matrices were generated for each fold and aggregated. Additionally, Receiver Operating Characteristic (ROC) curves were analyzed per class to assess the sensitivity and specificity of each model across SoC estimates. This evaluation framework provides a comprehensive comparison between a lightweight CNN architecture and a more complex deep model (AlexNet), offering insight into the trade-offs between accuracy, complexity, and training requirements in battery SoC estimation using GAF-encoded impedance data.

6.2 Results

This section presents the evaluation outcomes of two convolutional neural network architectures—Simple CNN and AlexNet—for estimating the State of Charge (SoC) of lithium-ion batteries using GAF-encoded impedance spectroscopy data. The models were trained and tested using a 10-fold cross-validation framework to ensure robustness and generalizability across different batteries.

6.2.1 Overall Model Performance

Table 6.6 summarizes the classification accuracy obtained by each model across the 10 folds of cross-validation. On average, the Simple CNN model demonstrated superior performance, achieving a mean accuracy of 77.75%, compared to 73.87% for AlexNet. This suggests that the relatively shallow architecture of Simple CNN may be more effective in learning generalized patterns from the GAF-transformed EIS data under the given conditions. Despite the overall advantage of Simple CNN, both models exhibited noticeable variability in accuracy across different folds. For example, AlexNet performed best in Fold-10 (88.75%) but showed a significant drop in Fold-8 (53.75%), indicating its sensitivity to certain battery-specific data patterns. Similarly, Simple CNN achieved its highest accuracy in Fold-10 (92.50%) and lowest in Fold-8 (62.50%).

Table 6.6: Classification accuracy (%) of Simple CNN and AlexNet across 10-fold cross-validation.

Fold	**AlexNet**	**Simple CNN**
Fold-1	86.25	80.00
Fold-2	72.50	76.25
Fold-3	72.50	77.50
Fold-4	78.75	76.25
Fold-5	78.75	80.00
Fold-6	76.25	82.50
Fold-7	65.00	82.50
Fold-8	53.75	62.50
Fold-9	66.25	67.50
Fold-10	88.75	92.50
Mean Accuracy	**73.87**	**77.75**

To assess the stability of the models across the 10-fold cross-validation, the mean accuracies and corresponding standard deviations are summarized in Table 6.7. This presentation provides additional insight into the consistency of the results, complementing the fold-wise accuracies shown in Table 6.6.

Table 6.7: Mean and Standard Deviation of Classification Accuracies for Battery SoC Estimation

Model	Mean Accuracy	Standard Deviation
Simple CNN	77.75%	8.64%
AlexNet	73.87%	10.43%

As observed in Table 6.7, the Simple CNN model achieves a higher mean accuracy compared to AlexNet, while also exhibiting a slightly lower standard deviation. This indicates that the Simple CNN not only performs better on average but is also more stable across folds.

The variability in performance underscores the challenges inherent in SoC estimation using impedance data. These include differences in battery aging profiles, non-uniform electrochemical behavior across SoC levels, and potential measurement noise. The relatively consistent performance of Simple CNN may be attributed to its simpler architecture, which reduces the risk of overfitting to noisy or battery-specific features.

6.2.2 Detailed Performance Metrics

Table 6.8 provides a detailed breakdown of the Simple CNN model's performance in terms of precision, recall, and F1-score for each fold. The model maintained a relatively stable classification profile across most folds, with an overall mean F1-score of 79.4%. Notably, the poorest performance was observed in Fold-8, where the F1-score dropped to 64%, further supporting the earlier observation that certain battery datasets introduce increased classification difficulty.

Table 6.8: Performance metrics (%) of Simple CNN across 10-fold cross-validation.

Fold	Precision	Recall	F1-Score
Fold-1	83	80	77
Fold-2	77	76	76
Fold-3	78	78	77
Fold-4	78	76	76
Fold-5	83	80	80
Fold-6	88	82	82
Fold-7	83	82	82
Fold-8	71	62	64
Fold-9	70	68	68
Fold-10	93	93	92
Mean	**80.4**	**79.7**	**79.4**

The fluctuation in F1-scores across folds reinforces the role of data variability in model performance. AlexNet, while generally less consistent than Simple CNN, achieved the highest single-fold accuracy (92.50%), indicating its potential to capture complex relationships in specific scenarios.

6.2.3 Analysis of Classification Behavior

The average confusion matrix, shown in Figure 6.6, offers further insights into the classification behavior of the Simple CNN model. The model performs well at distinguishing higher SoC levels such as SoC100, SoC90, and SoC80. However, classification errors become more frequent in the mid and lower SoC ranges. Misclassifications are especially prevalent between adjacent classes, such as SoC60 and SoC70 or SoC30 and SoC40, where impedance signatures tend to be less distinguishable. In some instances, misclassification occurs between non-adjacent classes (e.g., SoC90 misclassified as SoC10), which may be due to outliers or anomalous patterns not captured during training. These errors highlight a limitation of the current encoding and feature extraction approach in separating subtle variations in EIS characteristics, particularly in the low-to-mid SoC ranges.

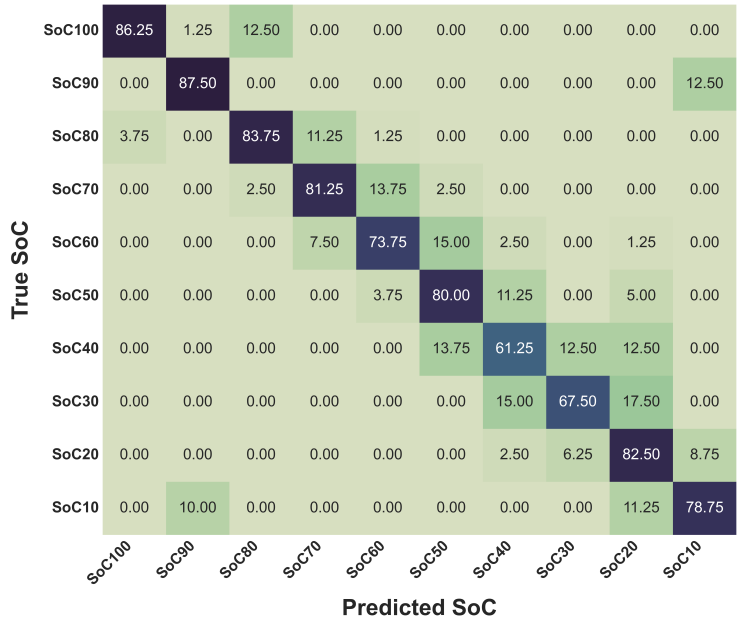


Figure 6.6: Mean confusion matrix for the Simple CNN model across 10 folds.

Overall, the results confirm that CNN-based models can effectively estimate SoC from GAF-encoded EIS data, although classification accuracy is influenced by the nonlinear and complex nature of impedance behavior across SoC levels. Future improvements may include more advanced feature encoding, the use of ensemble or hybrid models, and expansion of the dataset to better capture battery variability across usage conditions.

6.3 Summary

This chapter presented the process of generating, organizing, and utilizing an experimental dataset based on Electrochemical Impedance Spectroscopy (EIS) for the purpose of estimating the State of Charge (SoC) of lithium-based batteries. The dataset was obtained using a carefully designed measurement protocol and acquisition system, capturing high-resolution impedance spectra at discrete SoC levels across multiple charge-discharge cycles. A total of eleven cylindrical LFP batteries were tested under controlled thermal conditions, with EIS and capacity data stored in a structured and accessible format. The dataset includes both the real and imaginary parts of the impedance, recorded across a frequency sweep, and was transformed into image representations using the Gramian Angular Field (GAF) method. The image encoding step enabled the conversion of impedance spectra into two-dimensional formats suitable for deep learning applications. These encoded images preserved structural patterns in the EIS data and served as input to convolutional neural networks (CNNs) for classification-based SoC estimation. Two models—Simple CNN and AlexNet—were implemented and evaluated using a 10-fold cross-validation procedure. The evaluation strategy ensured that each battery contributed to training, validation, and testing in a rotation, allowing a robust and fair comparison of model generalization. Results indicated that the Simple CNN model offered higher average classification accuracy across folds (77.75%) and demonstrated more stable performance, while AlexNet achieved a higher peak performance in certain folds but suffered from greater variability. These outcomes highlight the trade-offs between model complexity and generalization when working with limited experimental data. Importantly, the limited dataset—comprising only two discharge-impedance cycles per battery—was found to be a key constraint. Although the structure of the dataset allowed for balanced class representation and systematic training, the small number of cycles limited the models' ability to capture broader electrochemical variability and led to reduced accuracy, especially at mid-to-low SoC levels. Given the data-intensive nature of deep learning methods, the expansion of this dataset is critical for enhancing model robustness and reliability.

Future work will focus on several directions: increasing the number of charge-discharge cycles per battery, expanding the range of battery chemistries and types, and incorporating more environmental conditions such as varying temperatures or degradation states. Additionally, improvements to the image encoding pipeline and exploration of advanced model architectures—such as attention-based CNNs, transformers, or physics-informed neural networks—may further enhance SoC estimation accuracy. The integration of uncertainty quantification could also provide greater confidence in model predictions, supporting practical deployment in battery management systems. Overall, this chapter has laid the foundation for impedance-based, image-driven battery SoC estimation using machine learning. It has demonstrated the feasibility of this approach while clearly identifying the challenges and opportunities that will guide subsequent research.

Conclusions and Future Works

6.4 Conclusions

This dissertation has presented a comprehensive study on the development and application of data-driven machine learning (ML) and deep learning (DL) methodologies for smart sensor systems, with particular emphasis on spectral color sensing, water pollutant detection, and battery State of Charge (SoC) estimation. The work leveraged the versatility of the SENSIPLUS platform to design integrated sensing solutions supported by advanced data preprocessing and model training pipelines.

6.4.1 Spectral Color Sensing with SENSIPATCH

The color-sensing experiments using the SENSIPATCH wearable device demonstrated that the system is capable of achieving high classification accuracy across a broad spectrum of PANTONE colors. In the preliminary study involving a limited color set, the Multilayer Perceptron (MLP) model achieved a mean accuracy of 91.63% under enclosed conditions (ENC), significantly outperforming other models such as KNN and Random Forest. The controlled black-box setup proved essential in enhancing data quality by minimizing ambient light interference.

In the extended experiments involving 100 PANTONE colors under varying contact conditions, the SVM model achieved the highest mean accuracy of 97.29% under firm contact conditions (FCCW), while the MLP provided the best generalization across combined contact scenarios with a mean accuracy of 96.08%. Notably, the results indicated that while firm contact improves classification performance due to stable sensor proximity, the system remained robust under loose contact, simulating realistic wearable deployment conditions.

The small performance degradation under loose contact conditions (SVM: 97.14%, MLP: 95.86%) highlights the practical viability of the SENSIPATCH system in dy-

namic environments, such as on human skin during movement. The combination of strong model performance across both controlled and variable contact scenarios underscores the suitability of this approach for real-time, wearable spectral sensing applications.

6.4.2 Water Pollutant Detection with SCW

In the water pollutant detection studies using the Smart Cable Water (SCW) system, deep learning models demonstrated strong capability in discriminating a diverse set of chemical pollutants using impedance spectroscopy data. In Experiment 1 involving eight pollutants, the CNN model achieved the highest classification performance, demonstrating its effectiveness in learning local signal patterns from sliding window segments.

Building on these results, Experiment 2 extended the pollutant set to 14 classes and introduced the Residual Channel and Spatial Attention Network (RCSANet). The RCSANet model significantly outperformed all baseline architectures, achieving a mean classification accuracy exceeding 94% across cross-validation folds. The integration of channel and spatial attention mechanisms allowed the model to dynamically focus on the most informative features and temporal regions, enhancing robustness and improving pollutant differentiation.

The consistent superiority of RCSANet over other models such as CNN, LSTM, and ResNet validates the effectiveness of attention-based architectures in processing multivariate impedance time-series data.

6.4.3 Battery SoC Estimation with EIS

The battery SoC estimation experiments using Electrochemical Impedance Spectroscopy (EIS) and image-based deep learning confirmed the feasibility of this approach for battery diagnostics. By transforming impedance data into Gramian Angular Field (GAF) images, the study enabled the application of Convolutional Neural Networks for SoC classification.

The Simple CNN model achieved an average classification accuracy of 77.75%, while AlexNet reached a maximum fold accuracy of 92.5%, albeit with higher performance variance. This highlights the potential of image-driven SoC estimation but also indicates the need for further research to improve stability, especially for mid-range SoC levels where class overlap remains a challenge.

6.4.4 Summary of Contributions

- Robust and scalable machine learning pipelines tailored for smart sensor signal classification.

- Effective preprocessing frameworks, including baseline correction, bootstrapping, and Z-score normalization to enhance sensor data quality.
- Comprehensive evaluation across three distinct sensing domains, demonstrating the generalizability of the proposed approaches.
- Development of the RCSANet architecture, achieving state-of-the-art pollutant classification performance in complex impedance datasets.
- Public release of a unique EIS dataset for battery SoC estimation to support future research.

Overall, the study validated the integration of smart sensor platforms with machine learning as a viable pathway to develop portable, low-cost, and accurate sensing solutions for real-time applications in health monitoring, environmental surveillance, and energy management. While the proposed frameworks achieved strong performance across color sensing, water pollutant detection, and battery SoC estimation, certain limitations should be acknowledged to provide a balanced perspective. The models remain sensitive to sensor drift, environmental noise, and unseen operating conditions, which may reduce their robustness in real-world deployments compared to controlled laboratory experiments. For instance, the battery SoC estimation task revealed greater misclassification in mid-range SoC levels, reflecting the inherent challenge of distinguishing subtle electrochemical variations. Addressing these limitations forms a key direction for future work. In particular, efforts will focus on developing noise-robust and uncertainty-aware learning strategies, expanding datasets to capture broader operational conditions, and employing domain adaptation techniques to mitigate device-specific drift. Such extensions will enhance model resilience and reliability, ensuring that the proposed methods can translate effectively from experimental validation to real-world applications.

6.5 Future Works

6.5.1 Machine Learning Methodologies

- **Transfer Learning and Domain Adaptation:** Future research could explore transfer learning to adapt pre-trained models to new sensor configurations, signal acquisition conditions, or different environmental settings. Domain adaptation techniques could also be employed to improve model generalization when working with heterogeneous data sources.
- **Online and Incremental Learning:** Implementing real-time adaptive learning strategies that allow models to update incrementally as new sensor data

become available would be valuable. This could enhance the system's ability to handle temporal drift, changing environmental factors, or evolving pollutant profiles without requiring full retraining.

- **Model Interpretability:** Incorporating explainable AI (XAI) frameworks to improve the interpretability and transparency of ML/DL models will be essential, particularly for applications like water quality assessment and battery diagnostics where model decisions must be trusted in practical deployment.

6.5.2 Data-Centric Improvements

- **Dataset Expansion:** Future studies should focus on expanding datasets across all applications. For water pollutant detection, larger datasets involving mixtures of pollutants, varying concentrations, and data collected from real-world environments would provide more challenging and practical classification scenarios. For SoC estimation, collecting additional impedance profiles under different battery types, usage conditions, and aging cycles would enhance model robustness.

- **Noise-Robust Learning:** Developing models that can maintain performance in the presence of noise, sensor drift, and environmental variability would further increase the reliability of data-driven smart sensing systems. Techniques such as data augmentation, adversarial training, and robust loss functions could be explored.

6.5.3 Advanced Deep Learning Techniques

- **Attention-Based Architectures:** Future research can investigate more advanced attention mechanisms or hybrid models that combine transformers with convolutional backbones to further enhance temporal and spatial feature extraction, particularly for complex sensing modalities such as multichannel impedance data.

- **Real-Time Inference Optimization:** Exploring lightweight deep learning models that can provide high accuracy while supporting real-time inference would make the systems more deployable. Techniques like model pruning, quantization, and knowledge distillation could be applied to reduce computational overhead.

- **Cross-Modal Learning:** Investigating the potential of fusing data from multiple sensor modalities (e.g., optical and electrochemical) using cross-

modal learning frameworks could yield more comprehensive and robust classification systems.

This future direction will continue to build on the data-driven foundation established in this thesis, focusing on improving model accuracy, generalizability, interpretability, and real-world applicability without requiring hardware modifications or redesign.

6.5.4 Dataset Expansion and Robustness

- Real-World Water Samples: Expanding the pollutant dataset to include complex mixtures, variable concentrations, and field-collected water samples to enhance model robustness and applicability in uncontrolled environments.
- Extended Battery Profiles: Augmenting the EIS dataset with additional battery types, operating conditions, and aging cycles to support more accurate SoC and battery health predictions.

6.5.5 Application-Specific Extensions

- Skin Diagnostics: Extending the spectral sensing methodology to dermatological applications such as non-invasive skin cancer screening, wound monitoring, and skin tone analysis.
- Smart Cities and IoT Networks: Deploying SCW sensors within distributed IoT frameworks for real-time environmental monitoring in urban and industrial areas.
- Battery Prognostics: Advancing the SoC framework toward predictive maintenance by incorporating battery health monitoring, cycle life forecasting, and anomaly detection capabilities.

This research has successfully demonstrated the synergistic potential of smart sensors and machine learning to address complex sensing challenges across multiple domains. The methodologies and architectures developed herein offer a strong foundation for future research and real-world deployment in next-generation wearable, environmental, and energy management systems.

Bibliography

[1] Chong-Min Kyung et al. “Smart sensors and systems”. In: *Cham, Switzerland: Springer* (2017).

[2] Manfred Glesner and François Philipp. “Embedded systems design for smart system integration.” In: *ISVLSI*. 2013, pp. 32–33.

[3] Domenico Formica and Emiliano Schena. *Smart sensors for healthcare and medical applications*. 2021.

[4] Qingxiao Tao et al. “Clinical applications of smart wearable sensors”. In: *Iscience* 26.9 (2023).

[5] T Lakshmi Narayana et al. “Advances in real time smart monitoring of environmental parameters using IoT and sensors”. In: *Heliyon* 10.7 (2024).

[6] Mohd Javaid et al. “Significance of sensors for industry 4.0: Roles, capabilities, and applications”. In: *Sensors International* 2 (2021), p. 100110.

[7] Atena Karbalaei Mohammad Hossein, Amir Golroo, and Medya Akhoundzadeh. “Smart embedded technologies and materials for enhanced pavement management”. In: *Automation in Construction* 168 (2024), p. 105865. ISSN: 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2024.105865. URL: https://www.sciencedirect.com/science/article/pii/S0926580524006010.

[8] Xiaozhou Ye et al. “Machine learning techniques for sensor-based human activity recognition with data heterogeneity—A review”. In: *Sensors (Basel, Switzerland)* 24.24 (2024), p. 7975.

[9] Weixiang Zhao et al. “Machine learning: A crucial tool for sensor design”. In: *Algorithms* 1.2 (2008), pp. 130–152.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.

[11] Jindong Wang et al. “Deep learning for sensor-based activity recognition: A survey”. In: *Pattern recognition letters* 119 (2019), pp. 3–11.

[12] SENSIPATCH. *https://sensichips.com/design-kit/wearables-kit/*. Accessed: 2025-04-13. 2025. URL: https://sensichips.com/design-kit/wearables-kit/.

[13] Hamza Mustafa et al. “A Deep Learning System for Water Pollutant Detection Based on the SENSIPLUS Microsensor”. In: *International Conference on Pattern Recognition*. Springer. 2024, pp. 192–203.

[14] Lewis D Griffin. “Optimality of the basic colour categories for classification”. In: *Journal of the Royal Society Interface* 3.6 (2006), pp. 71–85.

[15] Min Zhang et al. “Applications of color machine vision in the agricultural and food industries”. In: *Precision Agriculture and Biological Quality*. Ed. by George E. Meyer and James A. DeShazer. Vol. 3543. International Society for Optics and Photonics. SPIE, 1999, pp. 208–219. DOI: 10.1117/12.336908. URL: https://doi.org/10.1117/12.336908.

[16] Ming-Fang Cheng et al. “Modern Trends and Recent Applications of Hyperspectral Imaging: A Review”. In: *Technologies* 13.5 (2025). ISSN: 2227-7080. DOI: 10.3390/technologies13050170. URL: https://www.mdpi.com/2227-7080/13/5/170.

[17] Jiaojiao Zhou et al. “Recent Advances in Design and Application of Nanomaterials-Based Colorimetric Biosensors for Agri-food Safety Analysis”. In: *ACS Omega* 8.49 (2023), pp. 46346–46361. DOI: 10.1021/acsomega.3c06409. eprint: https://doi.org/10.1021/acsomega.3c06409. URL: https://doi.org/10.1021/acsomega.3c06409.

[18] Mansi Tripathi and Sunil Kumar Singal. “Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India”. In: *Ecological indicators* 96 (2019), pp. 430–436.

[19] Mehdi Mohammadi et al. “Deep learning for IoT big data and streaming analytics: A survey”. In: *IEEE Communications Surveys & Tutorials* 20.4 (2018), pp. 2923–2960.

[20] A Vaswani. “Attention is all you need”. In: *Advances in Neural Information Processing Systems* (2017).

[21] Andrea Ria et al. “The sensiplus: A single-chip fully programmable sensor interface”. In: *International Conference on Applications in Electronics Pervading Industry, Environment and Society*. Springer. 2021, pp. 256–261.

[22] Andrea Ria et al. “A Miniaturized ECG System Based on a Versatile Single Chip Sensor Interface”. In: *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE. 2024, pp. 1–4.

[23] M.A. Hannan et al. “A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations”. In: *Renewable and Sustainable Energy Reviews* 78 (2017), pp. 834–854. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2017.05.001. URL: https://www.sciencedirect.com/science/article/pii/S1364032117306275.

[24] Kong Soon Ng et al. "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries". In: *Applied Energy* 86.9 (2009), pp. 1506–1511. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2008.11.021. URL: https://www.sciencedirect.com/science/article/pii/S0306261908003061.

[25] Iryna Snihir et al. "Battery open-circuit voltage estimation by a method of statistical analysis". In: *Journal of Power Sources* 159.2 (2006), pp. 1484–1487.

[26] Xingtao Liu et al. "An extended Kalman filter based data-driven method for state of charge estimation of Li-ion batteries". In: *Journal of Energy Storage* 40 (2021), p. 102655.

[27] Hongfei Zhang, Zhumu Fu, and Fazhan Tao. "Improved sliding mode observer-based SOC estimation for lithium battery". In: *AIP Conference Proceedings*. Vol. 2122. 1. AIP Publishing. 2019.

[28] Jun Xu et al. "The state of charge estimation of lithium-ion batteries based on a proportional-integral observer". In: *IEEE Transactions on Vehicular Technology* 63.4 (2013), pp. 1614–1621.

[29] Rajakumar Sakile and Umesh Kumar Sinha. "Estimation of state of charge and state of health of lithium-ion batteries based on a new adaptive nonlinear observer". In: *Advanced Theory and Simulations* 4.11 (2021), p. 2100258.

[30] Ngoc-Tham Tran et al. "SOC estimation of multiple lithium-ion battery cells in a module using a nonlinear state observer and online parameter estimation". In: *Energies* 11.7 (2018), p. 1620.

[31] Ming Zhang et al. "Electrochemical impedance spectroscopy: A new chapter in the fast and accurate estimation of the state of health for lithium-ion batteries". In: *Energies* 16.4 (2023), p. 1599.

[32] Hamza Mustafa et al. "SoC estimation on Li-ion batteries: A new EIS-based dataset for data-driven applications". In: *Data in Brief* 57 (2024), p. 110947. ISSN: 2352-3409. DOI: https://doi.org/10.1016/j.dib.2024.110947. URL: https://www.sciencedirect.com/science/article/pii/S2352340924009107.

[33] Zhiguang Wang and Tim Oates. *Imaging Time-Series to Improve Classification and Imputation*. 2015. arXiv: 1506.00327 [cs.LG]. URL: https://arxiv.org/abs/1506.00327.

[34] M. Marsan, M. Lucidi, and G. Cincotti. "LED-Based Spectrophotometry". In: *2018 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*. 2018, pp. 1–2.

[35] M. Marsan, M. Lucidi, and G. Cincotti. "Geometrical-Optics Based Spectrophotometry". In: *2018 20th International Conference on Transparent Optical Networks (ICTON)*. 2018, pp. 1–1. DOI: 10.1109/ICTON.2018.8473846.

[36] C. Rosa et al. "Optomicrofluidic system for spectrophotometric analysis: automated process and wireless". In: *2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA)*. 2018, pp. 1–6. DOI: 10.1109/CCRA.2018.8588128.

[37] H. R. O. Rocha et al. "Machine Learning Method and Visible Light-Based Sensors for Multiinterface Level Measurement". In: *IEEE Sensors Journal* 23.14 (2023), pp. 16393–16401. DOI: 10.1109/JSEN.2023.3282026.

[38] C. Fragner et al. "LEDPOS: Indoor Visible Light Positioning Based on LED as Sensor and Machine Learning". In: *IEEE Access* 12 (2024), pp. 46444–46461. DOI: 10.1109/ACCESS.2024.3382757.

[39] A. Hina, H. Nadeem, and W. Saadeh. "A Single LED Photoplethysmography-Based Noninvasive Glucose Monitoring Prototype System". In: *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2019, pp. 1–5. DOI: 10.1109/ISCAS.2019.8702747.

[40] B. E. Manurung et al. "Non-Invasive Blood Glucose Monitoring using Near-Infrared Spectroscopy based on Internet of Things using Machine Learning". In: *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)*. 2019, pp. 5–11. DOI: 10.1109/R10-HTC47129.2019.9042479.

[41] D. M. Crafa et al. "Neural networks embedded in wearable devices: a preliminary digital vs. analog comparison". In: *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. 2023, pp. 587–592.

[42] M. d. P. B. Tobar et al. "Skinly: A novel handheld iot device for validating biophysical skin characteristics". In: *Skin Research and Technology* 30.3 (2024), e13613.

[43] L.-C. Hsu et al. "Developing low-cost mobile device and apps for accurate skin spectrum measurement via low-cost spectrum sensors and deep neural network technology". In: *Sensors* 22.22 (2022), p. 8844.

[44] M. A. Khan et al. "Internet of Things and Machine Learning–based Diabetes Management System". In: *2023 International Conference on IT and Industrial Technologies (ICIT)*. 2023, pp. 1–6. DOI: 10.1109/ICIT59216.2023.10335841.

[45] A. Zeb et al. "Learning Fruit Class from Short Wave Near Infrared Spectral Features, an AI Approach Towards Determining Fruit Type". In: *2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE)*. 2022, pp. 193–196. DOI: 10.1109/ICMRE54455.2022.9734107.

[46] S. B. Sulistyo et al. “A novel approach for detection of granulated coconut sugar adulteration using LED-based spectrometer and machine learning”. In: *Information Processing in Agriculture* (2024). DOI: 10.1016/j.inpa.2024.09.007.

[47] G. L. Hanopol and J. C. D. Cruz. “Rice Tungro Disease Detection Based on Light Absorption Assessment Using Machine Learning”. In: *2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC)*. 2022, pp. 145–150. DOI: 10.1109/ICSGRC55096.2022.9845158.

[48] S. Ueno and O. Sakai. “Data driven calibration of color-sensitive optical sensor by supervised learning for botanical application”. In: *IEEE Sensors Journal* 22.12 (2022), pp. 11915–11922.

[49] M. Greco et al. “A THz Imaging Scanner to Monitor Leaf Water Content”. In: *2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)* (2021), pp. 7–11. DOI: 10.1109/MetroAgriFor52389.2021.9628522.

[50] M. A. Tamayo-Monsalve et al. “Coffee Maturity Classification Using Convolutional Neural Networks and Transfer Learning”. In: *IEEE Access* 10 (2022), pp. 42971–42982. DOI: 10.1109/ACCESS.2022.3166515.

[51] C. Goyal et al. “Machine Learning-Based Low-Cost Colorimetric Sensor for pH and Free-Chlorine Measurement”. In: *IEEE Sensors Letters* 8.11 (2024), pp. 1–4. DOI: 10.1109/LSENS.2024.3473530.

[52] B. D. Batinic et al. “Using Machine Learning for Improvement of Reflected Spectrum Estimations of Colorimetric Probe”. In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–7. DOI: 10.1109/TIM.2020.3011763.

[53] K. Lee, I. Cho, and I. Park. “E-Nose System Based on Ultra-Low Power Single Micro-LED Gas Sensor and Deep Learning”. In: *2023 IEEE SENSORS*. 2023, pp. 1–4. DOI: 10.1109/SENSORS56945.2023.10324995.

[54] W. Zhu, P. Sang, and Y. He. “Facial skin colour classification using machine learning and hyperspectral imaging data”. In: *IET Image Processing* 16.2 (2022), pp. 509–520.

[55] Z. Li et al. “A data-driven approach for garment color classification in online fashion images”. In: *Electronic Imaging* 31 (2019), pp. 1–5.

[56] M. Garcia Garcia et al. “Graffiti identification system using low-cost sensors”. In: *To appear* (2024).

[57] J. Fan et al. “Machine-Learning Assisted Prediction of Spectral Power Distribution for Full-Spectrum White Light-Emitting Diode”. In: *IEEE Photonics Journal* 12.1 (2020), pp. 1–18. DOI: 10.1109/JPHOT.2019.2962818.

[58] M. Niu et al. "Visibility Constrained Wide-Band Illumination Spectrum Design for Seeing-in-the-Dark". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 13976–13985. DOI: 10.1109/CVPR52729.2023.01343.

[59] Y. Tang et al. "Active and low-cost hyperspectral imaging for the spectral analysis of a low-light environment". In: *Sensors* 23.3 (2023), p. 1437.

[60] Imran Shahid. "Carbon Based Sensors for Air Quality Monitoring Networks; Middle East Perspective". In: *Frontiers in Chemistry* (2024). DOI: 10.3389/fchem.2024.1391409.

[61] Yadvinder Singh et al. "Assessment of water quality condition and spatiotemporal patterns in selected wetlands of Punjab, India". In: *Environmental Science and Pollution Research* 29.2 (2022), pp. 2493–2509.

[62] Kofi Sarpong Adu-Manu et al. "Water Quality Monitoring Using Wireless Sensor Networks". In: *Acm Transactions on Sensor Networks* (2017). DOI: 10.1145/3005719.

[63] Yona Zakaria and Michael Kisangiri. "An Integrated Cloud-Based Wireless Sensor Network for Monitoring Industrial Wastewater Discharged Into Water Sources". In: *Wireless Sensor Network* (2017). DOI: 10.4236/wsn.2017.98016.

[64] Roberto Fernández Martínez et al. "On the Use of an IoT Integrated System for Water Quality Monitoring and Management in Wastewater Treatment Plants". In: *Water* (2020). DOI: 10.3390/w12041096.

[65] Yao Liu et al. "Recent progress regarding electrochemical sensors for the detection of typical pollutants in water environments". In: *Analytical Sciences* 38.1 (2022), pp. 55–70.

[66] Luca Gerevini et al. "An end-to-end real-time pollutants spilling recognition in wastewater based on the IoT-ready SENSIPLUS platform". In: *Journal of King Saud University - Computer and Information Sciences* 35.1 (2023), pp. 499–513. ISSN: 1319-1578. DOI: https://doi.org/10.1016/j.jksuci.2022.12.018. URL: https://www.sciencedirect.com/science/article/pii/S1319157822004360.

[67] Marco Ferdinandi et al. "A Novel Smart System for Contaminants Detection and Recognition in Water". In: *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*. 2019, pp. 186–191. DOI: 10.1109/SMARTCOMP.2019.00051.

[68] Mario Molinara et al. "An End to End Indoor Air Monitoring System Based on Machine Learning and SENSIPLUS Platform". In: *IEEE Access* 8 (2020), pp. 72204–72215. DOI: 10.1109/ACCESS.2020.2987756.

[69] C. Bourelly et al. “A Preliminary Solution for Anomaly Detection in Water Quality Monitoring”. In: *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*. 2020, pp. 410–415. DOI: 10.1109/SMARTCOMP50058.2020.00086.

[70] Alessandro Bria et al. “An IoT-ready solution for automated recognition of water contaminants”. In: *Pattern Recognition Letters* 135 (2020), pp. 188–195.

[71] Kevin Roitero et al. “Detection of Wastewater Pollution Through Natural Language Generation With a Low-Cost Sensing Platform”. In: *IEEE Access* 11 (2023), pp. 50272–50284. DOI: 10.1109/ACCESS.2023.3277535.

[72] Nida Nasir et al. “Water quality classification using machine learning algorithms”. In: *Journal of Water Process Engineering* 48 (2022), p. 102920. ISSN: 2214-7144. DOI: https://doi.org/10.1016/j.jwpe.2022.102920. URL: https://www.sciencedirect.com/science/article/pii/S2214714422003646.

[73] Víctor Flores, Ingrid Bravo, and Marcelo Saavedra. “Water Quality Classification and Machine Learning Model for Predicting Water Quality Statusmdash;A Study on Loa River Located in an Extremely Arid Environment: Atacama Desert”. In: *Water* 15.16 (2023). ISSN: 2073-4441. DOI: 10.3390/w15162868. URL: https://www.mdpi.com/2073-4441/15/16/2868.

[74] Smail Dilmi and Mohamed Ladjal. “A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques”. In: *Chemometrics and Intelligent Laboratory Systems* 214 (2021), p. 104329. ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2021.104329. URL: https://www.sciencedirect.com/science/article/pii/S0169743921000976.

[75] Shailesh Jaloree, Anil Rajput, and Sanjeev Gour. “Decision tree approach to build a model for water quality”. In: *Binary Journal of Data Mining & Networking* 4.1 (2014), pp. 25–28.

[76] Amir Hamzeh Haghiabi, Ali Heidar Nasrolahi, and Abbas Parsaie. “Water quality prediction using machine learning methods”. In: *Water Quality Research Journal* 53.1 (2018), pp. 3–13.

[77] Ahmad Ali AlZubi. “IoT-based automated water pollution treatment using machine learning classifiers”. In: *Environmental Technology* 45.12 (2024), pp. 2299–2307.

[78] Santosh Konde and Dr Shankar Deosarkar. “IOT based water quality monitoring system”. In: *2nd International Conference on Communication & Information Processing (ICCIP)*. 2020.

[79] Lingling Zhu et al. "Deep learning method for minimizing water pollution and air pollution in urban environment". In: *Urban Climate* 49 (2023), p. 101486. ISSN: 2212-0955. DOI: https://doi.org/10.1016/j.uclim.2023.101486. URL: https://www.sciencedirect.com/science/article/pii/S2212095523000809.

[80] Seung-Ho Kang, In-Seon Jeong, and Hyeong-Seok Lim. "A deep learning-based biomonitoring system for detecting water pollution using Caenorhabditis elegans swimming behaviors". In: *Ecological Informatics* 80 (2024), p. 102482. ISSN: 1574-9541. DOI: https://doi.org/10.1016/j.ecoinf.2024.102482. URL: https://www.sciencedirect.com/science/article/pii/S1574954124000244.

[81] G Charulatha et al. "Evaluation of ground water quality contaminants using linear regression and artificial neural network models". In: *Arabian Journal of Geosciences* 10 (2017), pp. 1–9.

[82] Cloé Desmet et al. "Electrochemical sensor for explosives precursors' detection in water". In: *Challenges* 8.1 (2017), p. 10.

[83] Scott N Dean et al. "Machine learning techniques for chemical identification using cyclic square wave voltammetry". In: *Sensors* 19.10 (2019), p. 2392.

[84] Kakoli Banerjee et al. "A machine-learning approach for prediction of water contamination using latitude, longitude, and elevation". In: *Water* 14.5 (2022), p. 728.

[85] Marco Ferdinandi et al. "A novel smart system for contaminants detection and recognition in water". In: *2019 IEEE international conference on smart computing (SMARTCOMP)*. IEEE. 2019, pp. 186–191.

[86] Luca Gerevini et al. "An end-to-end real-time pollutants spilling recognition in wastewater based on the IoT-ready SENSIPLUS platform". In: *Journal of King Saud University-Computer and Information Sciences* 35.1 (2023), pp. 499–513.

[87] Likun Xing, Liuyi Ling, and Xianyuan Wu. "Lithium-ion battery state-of-charge estimation based on a dual extended Kalman filter and BPNN correction". In: *Connection Science* 34.1 (2022), pp. 2332–2363. DOI: 10.1080/09540091.2022.2118675. eprint: https://doi.org/10.1080/09540091.2022.2118675. URL: https://doi.org/10.1080/09540091.2022.2118675.

[88] Hui Pang et al. "A Composite State of Charge Estimation for Electric Vehicle Lithium-Ion Batteries Using Back-Propagation Neural Network and Extended Kalman Particle Filter". In: *Journal of The Electrochemical Society* 169.11 (Nov. 2022), p. 110516. DOI: 10.1149/1945-7111/ac9f79. URL: https://dx.doi.org/10.1149/1945-7111/ac9f79.

[89] Gaoqi Lian et al. "Considering the temperature influence state-of-charge estimation for lithium-ion batteries based on a back propagation neural network and improved unscented Kalman filtering". In: *International Journal of Energy Research* 46.13 (2022), pp. 18192–18211.

[90] Jiashun Lin et al. "Algorithm of BPNN-UKF based on a fusion model for SOC estimation in lithium-ion batteries". In: *IET Power Electronics* 16.5 (2023), pp. 856–867.

[91] Chao Wang et al. "A two-stream network with complementary feature fusion for pest image classification". In: *Engineering Applications of Artificial Intelligence* 124 (2023), p. 106563. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2023.106563. URL: https://www.sciencedirect.com/science/article/pii/S0952197623007479.

[92] Molla S. Hossain Lipu et al. "Extreme Learning Machine Model for State-of-Charge Estimation of Lithium-Ion Battery Using Gravitational Search Algorithm". In: *IEEE Transactions on Industry Applications* 55.4 (2019), pp. 4225–4234. DOI: 10.1109/TIA.2019.2902532.

[93] Cheng Siong Chin and Zuchang Gao. "State-of-Charge Estimation of Battery Pack under Varying Ambient Temperature Using an Adaptive Sequential Extreme Learning Machine". In: *Energies* 11.4 (2018). ISSN: 1996-1073. DOI: 10.3390/en11040711. URL: https://www.mdpi.com/1996-1073/11/4/711.

[94] Jiaming Dou et al. "Extreme learning machine model for state-of-charge estimation of lithium-ion battery using salp swarm algorithm". In: *Journal of Energy Storage* 52 (2022), p. 104996. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2022.104996. URL: https://www.sciencedirect.com/science/article/pii/S2352152X22010015.

[95] Xiaobo Zhao et al. "State of charge estimation of lithium-ion battery based on multi-input extreme learning machine using online model parameter identification". In: *Journal of Energy Storage* 56 (2022), p. 105796. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2022.105796. URL: https://www.sciencedirect.com/science/article/pii/S2352152X22017844.

[96] Carlos Vidal et al. "Hybrid Energy Storage System State-Of-Charge Estimation Using Artificial Neural Network For Micro-Hybrid Applications". In: *2018 IEEE Transportation Electrification Conference and Expo (ITEC)*. 2018, pp. 1075–1081. DOI: 10.1109/ITEC.2018.8450251.

[97] Ran Li et al. "State of Charge Prediction Algorithm of Lithium-Ion Battery Based on PSO-SVR Cross Validation". In: *IEEE Access* 8 (2020), pp. 10234–10242. DOI: 10.1109/ACCESS.2020.2964852.

[98] Yuchen Song et al. "A hybrid statistical data-driven method for on-line joint state estimation of lithium-ion batteries". In: *Applied Energy* 261 (2020), p. 114408. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.114408. URL: https://www.sciencedirect.com/science/article/pii/S0306261919320951.

[99] Jiabo Li et al. "A Novel State of Charge Approach of Lithium Ion Battery Using Least Squares Support Vector Machine". In: *IEEE Access* 8 (2020), pp. 195398–195410. DOI: 10.1109/ACCESS.2020.3033451.

[100] Jinhao Meng, Guangzhao Luo, and Fei Gao. "Lithium Polymer Battery State-of-Charge Estimation Based on Adaptive Unscented Kalman Filter and Support Vector Machine". In: *IEEE Transactions on Power Electronics* 31.3 (2016), pp. 2226–2238. DOI: 10.1109/TPEL.2015.2439578.

[101] Junxiong Chen et al. "SOC estimation for lithium-ion battery using the LSTM-RNN with extended input and constrained output". In: *Energy* 262 (2023), p. 125375. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2022.125375. URL: https://www.sciencedirect.com/science/article/pii/S0360544222022575.

[102] Paul Takyi-Aninakwa et al. "An optimized relevant long short-term memory-squared gain extended Kalman filter for the state of charge estimation of lithium-ion batteries". In: *Energy* 260 (2022), p. 125093. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2022.125093. URL: https://www.sciencedirect.com/science/article/pii/S0360544222019880.

[103] Yong Chen et al. "SOC estimation of retired lithium-ion batteries for electric vehicle with improved particle filter by H-infinity filter". In: *Energy Reports* 9 (2023), pp. 1937–1947. ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2023.01.018. URL: https://www.sciencedirect.com/science/article/pii/S2352484723000203.

[104] Zuxin Li et al. "Novel method for modelling and adaptive estimation for SOC and SOH of lithium-ion batteries". In: *Journal of Energy Storage* 62 (2023), p. 106927. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2023.106927. URL: https://www.sciencedirect.com/science/article/pii/S2352152X23003249.

[105] Xiaofei Wang et al. "Noise immune state of charge estimation of li-ion battery via the extreme learning machine with mixture generalized maximum correntropy criterion". In: *Energy* 239 (2022), p. 122406. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2021.122406. URL: https://www.sciencedirect.com/science/article/pii/S0360544221026554.

[106] Chunsheng Hu et al. "Deep learning enabled state-of-charge estimation of LiFePO4 batteries: A systematic validation on state-of-the-art charging protocols". In: *Energy* 246 (2022), p. 123404. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2022.123404. URL: https://www.sciencedirect.com/science/article/pii/S0360544222003073.

[107] Chong Bian et al. "Robust state-of-charge estimation of Li-ion batteries based on multichannel convolutional and bidirectional recurrent neural networks". In: *Applied Soft Computing* 116 (2022), p. 108401. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2021.108401. URL: https://www.sciencedirect.com/science/article/pii/S1568494621011571.

[108] Biao Yang, Yinshuang Wang, and Yuedong Zhan. "Lithium Battery State-of-Charge Estimation Based on a Bayesian Optimization Bidirectional Long Short-Term Memory Neural Network". In: *Energies* 15.13 (2022). ISSN: 1996-1073. DOI: 10.3390/en15134670. URL: https://www.mdpi.com/1996-1073/15/13/4670.

[109] Jianlong Chen et al. "State-of-charge estimation of lithium-ion batteries using convolutional neural network with self-attention mechanism". In: *Journal of Electrochemical Energy Conversion and Storage* 20.3 (2023), p. 031010.

[110] Runmin Zou et al. "A novel convolutional informer network for deterministic and probabilistic state-of-charge estimation of lithium-ion batteries". In: *Journal of Energy Storage* 57 (2023), p. 106298. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2022.106298. URL: https://www.sciencedirect.com/science/article/pii/S2352152X22022873.

[111] M. S. Hossain Lipu et al. "Real-Time State of Charge Estimation of Lithium-Ion Batteries Using Optimized Random Forest Regression Algorithm". In: *IEEE Transactions on Intelligent Vehicles* 8.1 (2023), pp. 639–648. DOI: 10.1109/TIV.2022.3161301.

[112] Pritpal Singh et al. "Fuzzy logic modeling of EIS measurements on lithium-ion batteries". In: *Electrochimica Acta* 51.8 (2006). Electrochemical Impedance Spectroscopy, pp. 1673–1679. ISSN: 0013-4686. DOI: https://doi.org/10.1016/j.electacta.2005.02.143. URL: https://www.sciencedirect.com/science/article/pii/S0013468605007851.

[113] Jun Xu et al. "A new method to estimate the state of charge of lithium-ion batteries based on the battery impedance model". In: *Journal of Power Sources* 233 (2013), pp. 277–284. ISSN: 0378-7753. DOI: https://doi.org/10.1016/j.jpowsour.2013.01.094. URL: https://www.sciencedirect.com/science/article/pii/S0378775313001432.

[114] U. Westerhoff et al. “Electrochemical impedance spectroscopy based estimation of the state of charge of lithium-ion batteries”. In: *Journal of Energy Storage* 8 (2016), pp. 244–256. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2016.09.001. URL: https://www.sciencedirect.com/science/article/pii/S2352152X16301359.

[115] Marvin Messing et al. “Battery SoC Estimation from EIS using Neural Nets”. In: *2020 IEEE Transportation Electrification Conference Expo (ITEC)*. 2020, pp. 588–593. DOI: 10.1109/ITEC48692.2020.9161523.

[116] Iman Babaeiyazdi, Afshin Rezaei-Zare, and Shahab Shokrzadeh. “State of charge prediction of EV Li-ion batteries using EIS: A machine learning approach”. In: *Energy* 223 (2021), p. 120116. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2021.120116. URL: https://www.sciencedirect.com/science/article/pii/S0360544221003650.

[117] Emanuele Buchicchio et al. “Battery SOC estimation from EIS data based on machine learning and equivalent circuit model”. In: *Energy* 283 (2023), p. 128461. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2023.128461. URL: https://www.sciencedirect.com/science/article/pii/S0360544223018558.

[118] Carmine Bourelly et al. “EIS-Based SoC Estimation: A Novel Measurement Method for Optimizing Accuracy and Measurement Time”. In: *IEEE Access* 11 (2023), pp. 91472–91484. DOI: 10.1109/ACCESS.2023.3308029.

[119] Pantone. *https://www.pantone.com/*. Accessed: 2025-04-13. 2025. URL: https://www.pantone.com/.

[120] John Cristian Borges Gamboa. “Deep learning for time-series analysis”. In: *arXiv preprint arXiv:1701.01887* (2017).

[121] Alex Sherstinsky. “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”. In: *Physica D: Nonlinear Phenomena* 404 (2020), p. 132306.

[122] Zhiguang Wang, Weizhong Yan, and Tim Oates. “Time series classification from scratch with deep neural networks: A strong baseline”. In: *2017 International joint conference on neural networks (IJCNN)*. IEEE. 2017, pp. 1578–1585.

[123] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[124] GUM ISO. “ISO/IEC GUIDE 98-3: 2008, Guide to the expression of uncertainty in measurement”. In: *International Organisation for Standardisation, Geneva, Switzerland* (2008).

[125] Hoo-Chang Shin et al. “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1285–1298. DOI: 10.1109/TMI.2016.2528162.

[126] Hüseyin Eldem, Erkan Ülker, and Osman Yaşar Işıklı. “Alexnet architecture variations with transfer learning for classification of wound images”. In: *Engineering Science and Technology, an International Journal* 45 (2023), p. 101490. ISSN: 2215-0986. DOI: https://doi.org/10.1016/j.jestch.2023.101490. URL: https://www.sciencedirect.com/science/article/pii/S2215098623001684.