Title

Machine Learning Based Multimodal Sensor Data Analysis for Nucleate Boiling Heat Transfer Assessment on ZnO Nanostructured Surfaces

ISBN

9798189270659

Author

Tchouteng Njike, Ursan

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Machine Learning Based Multimodal Sensor Data Analysis for Nucleate Boiling Heat Transfer Assessment on ZnO Nanostructured Surfaces

by

Ursan Tchouteng Njike

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Van P. Carey, Chair
Professor Tarek Zohdi
Professor Thomas Schutzius

Spring 2026

Machine Learning Based Multimodal Sensor Data Analysis for Nucleate Boiling Heat Transfer Assessment on ZnO Nanostructured Surfaces

By

Ursan Tchouteng Njike

Abstract

Machine Learning Based Multimodal Sensor Data Analysis for Nucleate Boiling Heat Transfer Assessment on ZnO Nanostructured Surfaces

by
Ursan Tchouteng Njike
Doctor of Philosophy in Mechanical Engineering
University of California, Berkeley
Professor Van P. Carey, Chair

This dissertation presents a systematic investigation of machine learning strategies for analyzing boiling heat transfer on zinc oxide (ZnO) nanostructured surfaces, progressing from single-modality digital data analysis to multimodal sensor fusion incorporating image, digital, and acoustic data. ZnO nanopillar surfaces with ultra-low contact angles below 10 degrees exhibit distinct vaporization regimes during water droplet deposition experiments at surface superheats ranging from 10°C to 55°C.

The research advances through three phases. First, three convolutional neural network (CNN) architectures with progressively more complex input modalities are developed for combined image and sensor data analysis. A digital-only baseline (Case A) achieves 15.8% root mean squared percent error (RMSPE) in mean heat flux prediction. A hybrid parallel-series CNN incorporating high-speed video images with non-thermal digital data (Case B) reduces this to 10.3% RMSPE, demonstrating that the CNN effectively functions as a non-contact temperature sensor by extracting thermal state information from visual morphology without direct temperature measurement. Adding all available digital inputs through a skip connection architecture (Case C) achieves 5.8% RMSPE and 96.9% regime classification accuracy.

Second, acoustic emission monitoring is introduced as an additional sensing modality. A neural network trained on quartile-based acoustic features (mean absolute amplitude, mean frequency, and frequency standard deviation computed over four equal temporal segments of each signal) achieves 13.3% RMSPE in predicting surface superheat from sound alone, demonstrating that acoustic emissions carry quantitative information about bubble nucleation and collapse dynamics complementary to visual and thermal measurements.

Third, a multimodal fusion network incorporating image, digital, and acoustic data achieves 4.6% RMSPE in heat flux prediction and 100% boiling regime classification accuracy. Permutation importance analysis confirms that image and acoustic features contribute most, with acoustic information providing its greatest advantage in the vigorous nucleation and Partial Dryout Vapor Recoil (PDVR) regimes.

The progressive improvement from 15.8% to 4.6% RMSPE provides compelling evidence that multimodal sensor fusion represents a powerful paradigm for analyzing complex thermal-fluid systems, with broad applicability to flow boiling, condensation, and nuclear thermal-hydraulics.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Van P. Carey, for his patience, guidance, and unwavering support throughout my doctoral studies. His expertise in two-phase heat transfer and his encouragement to explore machine learning approaches shaped the direction of this research from its inception. Professor Carey's ability to see connections between classical heat transfer physics and modern data-driven methods gave this work its distinctive character, and his mentorship extended well beyond the technical.

I wish to thank Professor Tarek Zohdi, who served on both my qualifying examination committee and my dissertation committee. His course on Modeling and Simulation of Advanced Manufacturing Processes shaped my understanding of how computational methods can augment physical intuition in engineering analysis. I am also grateful to Professor Thomas Schutzius, who chaired my qualifying examination committee and served on my dissertation committee. His expertise in surface engineering and interfacial transport phenomena provided valuable perspective on the nanostructured surfaces central to this work, and his feedback strengthened the rigor of this dissertation.

I am grateful to my colleagues in the Energy and Multiphase Transport Laboratory. Anisa Silva deserves special recognition for her collaboration on the experimental work and published papers that form the foundation of this dissertation; her contributions to the data collection and experimental campaigns were indispensable. I also thank Emma McClure and Samuel Cabrera, whose earlier investigations of ZnO nanostructured surfaces established the experimental platform and surface fabrication protocols used throughout this research. Their foundational work on surface characterization and droplet vaporization experiments made the machine learning investigations presented here possible.

To my family — thank you for your encouragement, sacrifice, and unwavering belief in me throughout this journey. Your support and presence made the path, though challenging, one I will always cherish.

This work was supported by the National Science Foundation through the Graduate Research Fellowship Program and NSF CBET grant number 2228373. Funding for this research was also provided by the endowment of the A. Richard Newton Chair in Engineering at UC Berkeley.

Chapter 1.
ZnO Nanostructured Surfaces and Boiling Heat Transfer Background

1.1. Introduction

Boiling heat transfer is one of the most effective mechanisms for thermal management in a wide range of engineering systems, with applications ranging from power generation and electronics cooling to chemical processing and nuclear reactor safety. The interest in enhancing boiling performance has driven decades of research into surface modification strategies that can increase nucleation site density, improve liquid supply to the heated surface, and delay the onset of critical heat flux (CHF) conditions. Among the many surface engineering approaches explored, nanostructured coatings have emerged as particularly promising due to their ability to fundamentally alter the wetting, wicking, and nucleation characteristics of heated surfaces [1,2].

The need for improved boiling heat transfer performance continues to grow with the increasing power densities of modern electronic devices, the push toward more compact and efficient heat exchangers in power generation systems, and the safety requirements of nuclear reactor thermal-hydraulic systems. Traditional approaches to enhancing boiling performance have included surface roughening, the creation of reentrant cavities, the application of microporous coatings, and the use of structured surfaces with well-defined geometries. While these approaches have achieved various degrees of success, the emergence of nanostructured surfaces has opened new possibilities by enabling surface modification at the scales comparable to the initial stages of bubble nucleation.

This dissertation focuses on the development and application of machine learning strategies for analyzing boiling heat transfer on zinc oxide (ZnO) nanostructured surfaces. These surfaces, which consist of dense arrays of ZnO nanopillars grown on metallic substrates via hydrothermal synthesis, exhibit superhydrophilic behavior with contact angles below 10 degrees and extreme wicking capabilities [1,2]. These properties lead to rapid droplet spreading upon contact with the surface and a fundamentally different vaporization behavior compared to uncoated surfaces. When water droplets are deposited on superheated ZnO nanostructured surfaces, the resulting vaporization process progresses through distinct regimes as the surface superheat increases, analogous to the well-known regimes in pool boiling but with important differences arising from the nanostructured surface properties.

Building on prior work in our research group that employed genetic algorithms for analysis of pool boiling quenching data [3], the research presented in this dissertation systematically progresses through increasingly sophisticated convolutional neural network (CNN) architectures for combined image and digital sensor data analysis, culminating in a multimodal sensor fusion framework that incorporates acoustic emission data alongside imaging and traditional sensor measurements. This progression represents a methodical exploration of how machine learning can enhance our understanding and prediction of boiling heat transfer phenomena on nanostructured surfaces. By combining acoustic monitoring, high-speed videography, and optimal measurement techniques with digital sensors, this work demonstrates that a robust suite of experimental tools, when analyzed through appropriate machine learning architectures, can yield significantly improved predictions of heat transfer performance.

1.1. ZnO Nanopillar Surface Fabrication

The nanostructured surfaces used throughout this dissertation were fabricated using a hydrothermal synthesis process to grow ZnO nanopillars on metallic substrates. The fabrication procedure, which has been refined through multiple investigations in our research group [1,2,4], involves several key steps that are described in detail below.

First, the metallic substrate (copper in this study) is machined to the desired geometry, a 2.5 cm diameter cylinder, and thoroughly cleaned via sonication in acetone, isopropyl alcohol, and distilled water. Non-growth surfaces are masked with Teflon tape to prevent contamination during the growth process. The substrate is then drilled on the side, approximately 1 mm from the top surface, with two holes positioned opposite to each other for the insertion of K-type thermocouples for surface temperature measurements during subsequent experiments.

A growth solution is prepared from hexamethylene tetramine (HMT, $C_6H_{12}N_4$) and zinc nitrate hexahydrate ($Zn(NO_3)_2{\cdot}6H_2O$) at equimolar concentration of 0.25 mol/L. The solution is stirred for two hours to ensure complete dissolution and uniform mixing. Prior to immersion in the growth solution, the copper surface is seeded with ZnO nanoparticles by depositing multiple layers of a diluted ZnO nanoparticle solution (approximately 6 nm particle size), followed by annealing at 120°C to promote adhesion and create nucleation sites for subsequent nanopillar growth [1,4]. The seeded substrate is then immersed in the growth solution and placed in an oven at approximately 90°C for a controlled growth period. The growth time can be varied to control the resulting nanostructure morphology, with typical growth periods ranging from 4 to 24 hours.

The resulting nanostructure, shown in Figure 1.1, consists of dense arrays of ZnO nanopillars with typical dimensions of approximately 60 nm cross-section, 50 nm spacing between pillars, and 400 nm pillar length, forming a layer approximately 5 μm thick [1,2]. These structural features create a nanoporous morphology that exhibits extreme wetting behavior, with water contact angles typically in the range of 6–10 degrees, which classifies the surface as superhydrophilic. The ultra-low contact angle and nanoscale porosity combine to produce exceptional wicking characteristics, enabling rapid lateral transport of liquid across the surface. The ZnO nanopillars adopt the wurtzite crystal structure (hexagonal), which is the thermodynamically stable phase of ZnO at ambient conditions. The pillar morphology is stochastically arranged, reflecting the randomness of the nucleation sites on the seeded substrate.

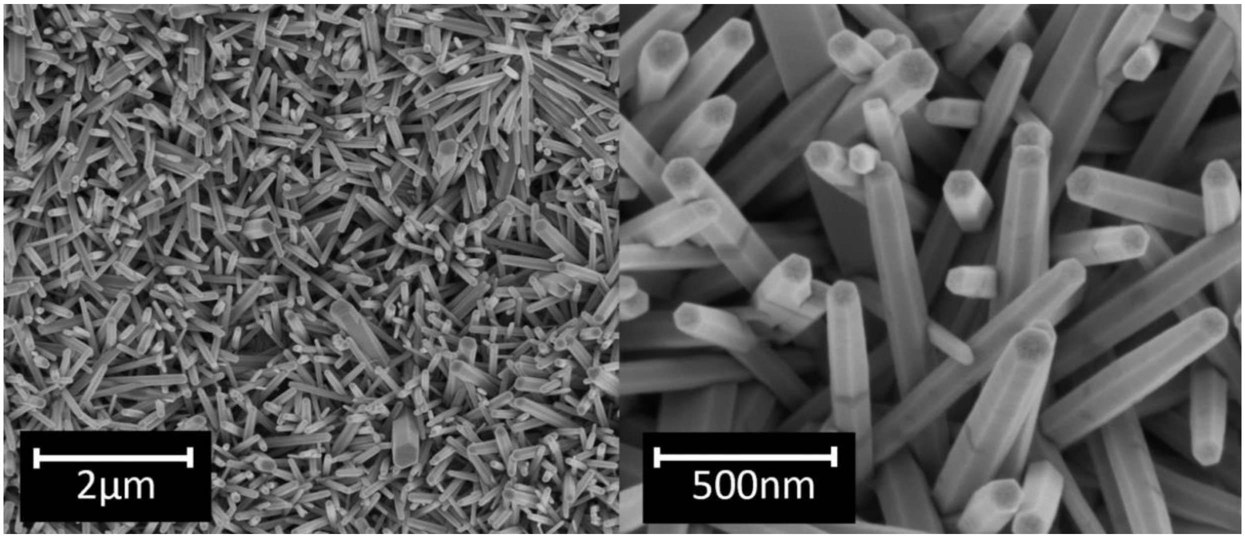
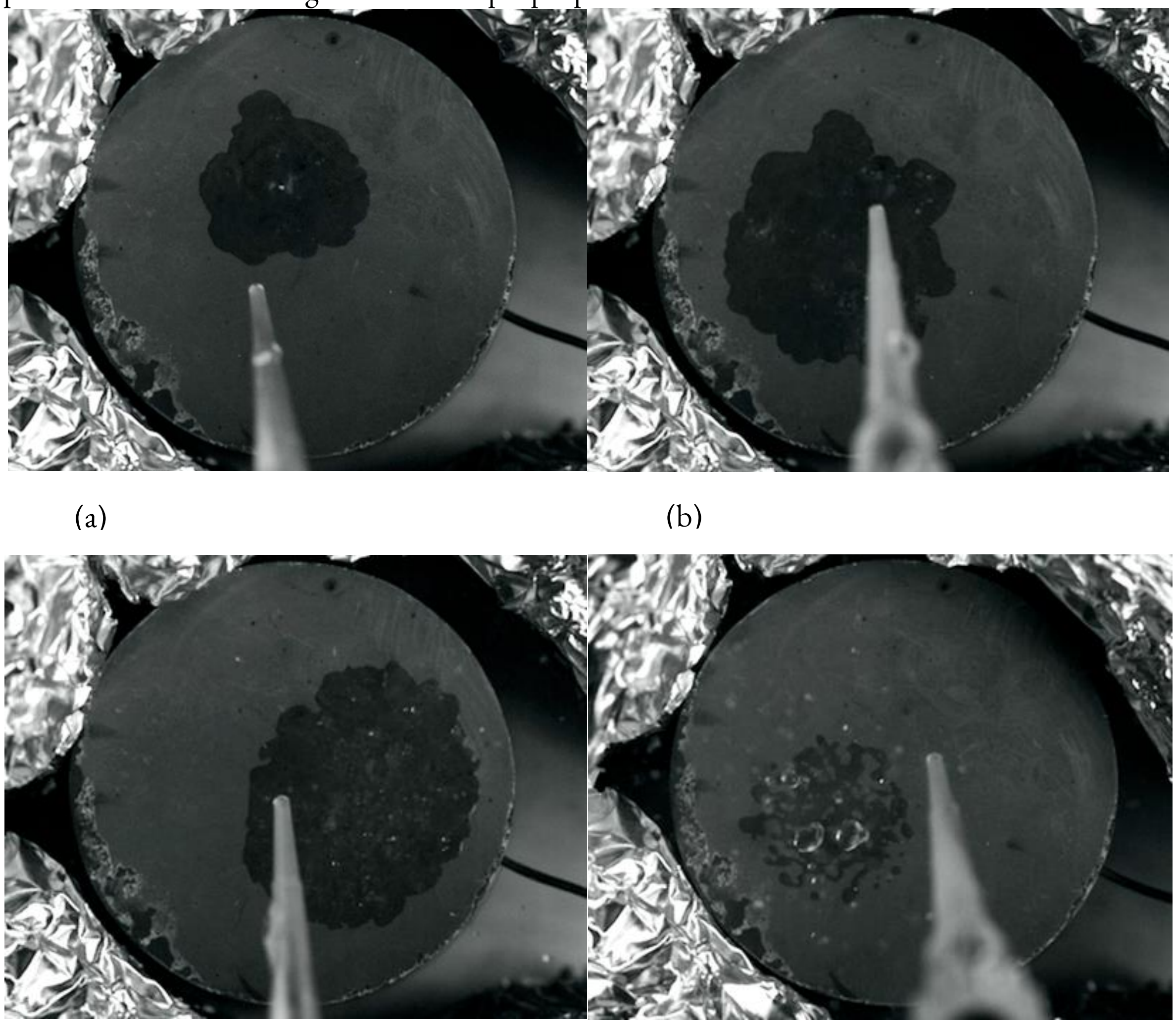


Figure 1.1. SEM micrograph of ZnO nanostructured surface showing nanopillar morphology

The contact angles of water on the coated surfaces were determined using adiabatic surface footprint measurements and a spherical cap model for the droplet geometry. For a droplet of known volume V_0 with measured footprint radius R, the apparent contact angle is computed from the spherical cap relationship. The footprint radius was measured from high-speed video frames and the droplet volume was known from the pipette calibration, allowing the apparent contact angle to be determined for each surface [2]. Previous investigations in our research group have established that the hydrothermal growth process is repeatable and produces surfaces with consistent wetting properties when the growth parameters (concentration, temperature, time, and seeding procedure) are carefully controlled [4].
The superhydrophilic nature of these surfaces has important consequences for the boiling heat transfer process. When a water droplet is deposited on the surface, liquid quickly fills the nanostructure due to the high capillary pressure resulting from the nanoscale spacing between pillars. This capillary pressure also makes the surface "sticky," preventing droplets from bouncing off, which reduces the volume of water wasted in cooling applications. The droplet spreading on the surface exhibits two distinct regimes: a synchronous spreading regime where the liquid front in the nanostructure and the upper droplet contact line spread together, followed by a hemi-spreading regime where the upper droplet settles at the equilibrium contact angle while liquid continues to spread through only the nanostructure, creating a characteristic "fried egg" appearance [4]. An 8 μL droplet typically spreads to a mean thickness of approximately 70 μm within the first 50 ms after deposition, creating a thin liquid film with a large wetted footprint that promotes efficient heat transfer.

1.2. Boiling Regimes on Nanostructured Surfaces

Previous investigations of pool boiling and droplet vaporization on ZnO nanostructured surfaces have revealed that the vaporization process exhibits distinct regimes that are analogous to those observed in pool boiling on conventional surfaces [4,5]. The behavior of water droplets deposited on superheated nanostructured surfaces progresses through four identifiable regimes as the surface superheat increases, as illustrated in Figure 1.2. These regimes correspond to the well-known regimes in pool boiling, conduction-dominated evaporation, nucleate boiling, transition boiling, and film boiling, but with important differences arising from the unique properties of the nanostructured surface.



(a) (b) (c) (d)

Figure 1.2. Representative images of vaporization regimes: (a) conduction, (b) isolated bubble, (c) vigorous nucleation, (d) PDVR

At low superheat ($\Delta T < 15°C$), the dominant heat transfer mechanism is conduction through the thin liquid layer with gentle evaporation at the liquid–air interface. In this regime, the droplet maintains a stable, thin, spread morphology due to the superhydrophilic nature of the surface. The liquid film remains largely quiescent, with heat transfer occurring primarily through conduction from the heated surface through the thin liquid film to the evaporating liquid–air interface. Because the nanostructured surface promotes extreme spreading, the liquid film is very thin relative to its footprint area, which results in relatively high heat transfer rates even in this conduction-dominated regime compared to uncoated surfaces.

At moderate superheat (15–25°C), isolated bubbles begin to nucleate at active sites within the spread droplet. These bubbles grow, depart, and are replaced by liquid drawn in by the wicking action of the nanostructure, representing the onset of nucleate boiling (ONB). Research presented by Wemp [1] found similar results, observing no nucleate boiling at 15°C superheat and bubble nucleation at 20°C, indicating the onset of nucleate boiling occurs between these two superheat values. The isolated bubble regime is characterized by the periodic appearance and departure of individual bubbles at discrete nucleation sites, with the liquid film between sites remaining relatively undisturbed. The wicking capability of the nanostructured surface plays a critical role in this regime by rapidly rewetting the surface after bubble departure, maintaining contact between liquid and the heated surface.

As superheat increases further (25–40°C), vigorous nucleation occurs with multiple bubbles simultaneously active across the droplet footprint. In this regime, bubble interactions including coalescence and departure create a complex, dynamic two-phase morphology that changes rapidly with time. The high density of active nucleation sites leads to significant bubble–bubble interaction, including lateral coalescence where adjacent bubbles merge before departure, and vertical coalescence where departing bubbles absorb newly forming bubbles in their wake. The morphology in this regime is significantly more complex than the isolated bubble regime, presenting a challenging test case for machine learning analysis.

At the highest superheats (40–55°C), the process enters the Partial Dryout Vapor Recoil (PDVR) regime, characterized by two mechanisms not present at lower superheats [5,6]: (1) local dryout in the spread droplet due to rapid vaporization at nucleation sites, where the rate of vaporization locally exceeds the rate at which liquid can be supplied by wicking, and (2) vapor recoil forces that eject liquid away from the surface. The vapor recoil phenomenon occurs when the rapid vaporization of liquid at the heated surface creates a high-velocity vapor flow that exerts a reaction force on the liquid, physically pushing it away from the surface. In this regime, the vaporization process becomes chaotic, alternating between rapid vaporization of liquid in contact with the surface and ejection of liquid off the surface by strong vapor recoil forces. The two-phase morphology is highly complex and changes rapidly, making this regime particularly challenging for both experimental observation and machine learning analysis.

Each of these regimes produces distinct visual appearances in high-speed video imagery and, as demonstrated in this dissertation, distinct acoustic signatures that correlate with the underlying heat transfer mechanisms. The systematic identification and characterization of these regimes provides the foundation for the machine learning approaches developed in subsequent chapters.

1.3. Machine Learning in Heat Transfer Research

The application of machine learning to heat transfer research has grown substantially in recent years, driven by the increasing availability of high-quality experimental data and advances in computational capabilities. Several categories of machine learning tools have been applied to boiling heat transfer problems, each offering distinct advantages depending on the nature of the data and the research objectives.

Genetic algorithms (GAs) represent a class of optimization techniques inspired by natural selection. In the context of heat transfer research, GAs have been used to fit physics-inspired model equations to experimental data, allowing researchers to explore the parametric dependencies of heat transfer performance on operating conditions and surface properties. McClure and Carey [7] applied genetic algorithms combined with deep learning to explore parametric trends in nucleate boiling heat transfer data, demonstrating that GAs can effectively identify the functional form and coefficient values of correlating equations that capture the essential physics of the boiling process. Their work showed that GA-based optimization could identify models that account for the effects of wall superheat, gravity, Marangoni effects, and pressure for binary mixtures. In my initial research activities [3], GAs were used to assess surface dryout during nucleate and transition boiling on surfaces with different wetting and substrate properties. That study analyzed quenching pool boiling data across the full boiling curve, from nucleate boiling through maximum heat flux to transition boiling and the Leidenfrost point, using genetic algorithms combined with heuristic models of dryout parametric dependence. The central hypothesis was that the boiling sequence from nucleate through transition boiling is a consequence of progressive surface dryout as wall superheat increases, and the GA analysis confirmed that substrate material properties have a significant effect on progressive surface dryout. Silva and Carey [8] further applied genetic algorithms to model the interaction of conduction and nucleate boiling mechanisms during evaporation of water droplets on superheated ZnO nanostructured surfaces.

Artificial neural networks (ANNs) have been applied to predict boiling heat transfer coefficients from digital sensor data by learning nonlinear relationships between input parameters (such as pressure, mass flux, quality, and surface properties) and output heat transfer quantities. Sajjad et al. [9] developed a high-fidelity ANN approach to correlate nucleate pool boiling data of roughened surfaces. Qiu et al. [10] created an artificial neural network model to predict mini/micro-channel saturated flow boiling heat transfer coefficients based on universal consolidated data. Calati et al. [11] applied ANNs to predict water pool boiling in metal foams, developing a generalized model from experimental results. Kuang et al. [12] investigated saturated hydrogen nucleate flow boiling heat transfer coefficients using neural networks. Earlier foundational work by Swain and Das [13] and Kishor and Das [14] demonstrated the feasibility of using soft computing techniques for predicting boiling heat transfer coefficients on copper-coated tubes, establishing the potential of neural network approaches for boiling heat transfer correlation.

More recently, CNNs have enabled direct analysis of image data from boiling experiments through learnable convolution filters that can extract spatial features from images without manual feature engineering. The paper of Yang et al. [15] discusses the use of high-speed camera visualization to examine flow boiling in microchannels and use a CNN model to identify flow patterns for use as a means of accurate heat transfer mechanism prediction. Lee [16] explored strategies to optimize experimental observation and measurements of pool boiling heat transfer using computer vision techniques, performing bubble detection and segmentation to track bubble movements as they depart from the surface, autonomously capturing key performance parameters such as bubble density, size, and departure frequency. Most commonly, CNN analysis of boiling process images has been done in parallel with separate analysis to fit physics-based or machine learning models to the digital data from the experiment. Because past research has established a correlation between superheat level and boiling regimes, separate CNN models predicting superheat from image features have been developed successfully. However, training a CNN model that simultaneously analyzes both digital data and images, as developed in this dissertation, represents a more powerful approach that can capture the complementary information content of both modalities.

The comprehensive review by Carey [17] provides an extensive landscape assessment of machine learning applications in heat transfer research, documenting the evolution from artificial neural network approaches through genetic algorithms to modern deep learning architectures. This review establishes the context for the multimodal approaches developed in this dissertation and explicitly identifies real-time multimodal processing, the ability to swiftly process optical images, thermal data, and acoustics, as opening new frontiers for heat transfer research.

1.4. Scope and Organization of the Dissertation

This dissertation presents a systematic investigation of machine learning strategies for analyzing boiling heat transfer on ZnO nanostructured surfaces, progressing from single-modality digital data analysis to multimodal sensor fusion incorporating image, digital, and acoustic data. The work builds on four published or submitted research papers that represent the progressive development of these techniques.

This investigation is guided by three central hypotheses, each tested through the progressive development of CNN architectures presented in Chapters 3 and 4:

Hypothesis 1: High-speed video images of the boiling process contain quantitative information about heat transfer performance that is complementary to, and not fully captured by, traditional digital sensor measurements (temperature, droplet volume, contact angle). If true, a CNN model trained on combined image and digital data should achieve significantly lower prediction error than a model trained on digital data alone, and the CNN should be able to infer thermal state variables (such as surface superheat) from visual morphology without direct temperature input.

Hypothesis 2: Acoustic emissions generated during droplet vaporization carry information about bubble dynamics — nucleation frequency, growth rate, collapse intensity, and temporal evolution — that is complementary to both visual and digital data. If true, incorporating acoustic features into the multimodal framework should yield measurable improvement in prediction accuracy beyond what image-plus-digital models achieve, and the acoustic-only model should demonstrate non-trivial predictive capability for thermal state variables.

Hypothesis 3: The prediction accuracy of the multimodal fusion model approaches the fundamental limits imposed by experimental measurement uncertainty and the inherent stochasticity of the boiling process. If true, the residual prediction error of the most comprehensive model should be comparable in magnitude to the estimated experimental uncertainty in the heat flux measurement, suggesting that the combined sensor suite captures nearly all deterministic information about the boiling process.

These hypotheses are tested quantitatively through the RMSPE metric applied to heat flux prediction and through classification accuracy applied to regime identification, with each successive architecture (Cases A, B, C, and the multimodal fusion network) designed to isolate the contribution of a specific data modality.

The dissertation is organized into seven chapters. Chapter 1 (this chapter) provides the background on ZnO nanostructured surfaces, boiling regimes, and machine learning approaches relevant to this research. It also explains why morphology imaging and acoustic characterization connect to heat transfer mechanisms: the visual appearance of the two-phase mixture on the surface directly reflects the active heat transfer mechanisms, while acoustic emissions carry information about bubble nucleation, growth, coalescence, and collapse dynamics that is complementary to visual observations. The characteristic sounds of boiling, ranging from the gentle crackling of incipient nucleation to the violent popping associated with vapor recoil events, are direct manifestations of bubble dynamics. Each boiling regime produces distinct acoustic signatures because the frequency, amplitude, and temporal pattern of acoustic emissions are governed by the size, growth rate, and collapse dynamics of bubbles, all of which vary systematically with superheat.

Chapter 2 reviews prior work on multimodal information processing across diverse technology domains, establishing the broader context for the multimodal approaches developed in this dissertation. This survey covers medical imaging, autonomous vehicles, industrial process monitoring, environmental monitoring, audio-visual processing, and structural health monitoring, identifying cross-domain principles that inform the architecture choices made in Chapters 3 and 4.

Chapter 3 presents the development of CNN-based approaches for combined image and sensor data analysis, based on our publications in the ASME Heat Transfer Summer Conference [18] and the ASME Journal of Heat and Mass Transfer [5]. This chapter details three progressively more sophisticated CNN architectures (Cases A, B, and C) and demonstrates the substantial improvement achieved by incorporating image data alongside traditional digital measurements.

Chapter 4 extends the multimodal framework to include acoustic emission data, based on our submission to the ASME 2026 Heat Transfer Summer Conference [19]. This chapter presents the acoustic signal characteristics, the novel quartile-based feature extraction approach, the acoustic-only neural network, and the multimodal fusion network that achieves the best prediction performance.

Chapter 5 provides an assessment of how multimodal information analysis enhances boiling research, synthesizing the progressive improvements demonstrated across the preceding chapters and discussing the physical insights gained from the multimodal approach.

Chapter 6 explores future directions for multimodal modeling, examining the potential to enhance heat transfer and energy research across multiple configurations (flow boiling, condensation, nuclear thermal-hydraulics, data center cooling) as well as contributions to the development of adaptive and autonomous technologies (digital twins, smart manufacturing, real-time process control, edge deployment).

Chapter 7 presents the concluding remarks, summarizing the key contributions of this dissertation, discussing their implications for the broader field of boiling heat transfer research, and offering recommendations for future work that could extend the multimodal framework to new configurations, sensing modalities, and application domains.

Chapter 2.
Prior work on multimodal information processing development for other technologies and systems

The multimodal sensor fusion approach developed in this dissertation for boiling heat transfer analysis has significant precedent across diverse technology domains. Multimodal machine learning, the fusion of images, audio, sensor time-series, and other data modalities through deep neural networks, has become a dominant paradigm in fields ranging from medical diagnostics to autonomous driving. Across all domains surveyed in this chapter, multimodal approaches consistently outperform single-modality methods by 5–20% in accuracy or equivalent metrics, with the advantage amplifying under degraded conditions [20]. The architectural landscape has evolved from simple feature concatenation toward attention-based cross-modal fusion and transformer architectures, while the frontier now includes self-supervised foundation models for heterogeneous sensor data.

This chapter reviews the most relevant prior work in six application domains to establish context for the contributions of this dissertation. Each domain is examined with particular attention to three aspects: the sensing modalities that are combined and what complementary physical information each provides; the neural network architectures used for fusion and why they outperform single-modality approaches; and the quantitative improvement achieved through multimodal fusion, with emphasis on results that parallel or inform the boiling heat transfer application. The chapter concludes with a synthesis of cross-domain principles that directly guided the architecture and methodology choices described in Chapters 3 and 4.

2.1. Medical Imaging and Diagnostics

Multimodal fusion has achieved some of its most mature results in oncology and neurology, where combining imaging with genomic, pathological, and clinical data has produced measurable improvements in diagnosis and prognosis. The medical domain is relevant to this dissertation because it demonstrates the core principle underlying all multimodal approaches: different sensing modalities capture fundamentally different physical information about the same system, and their fusion captures relationships that are invisible to any single modality.

Chen et al. [21] developed PORPOISE, a system fusing whole slide images (WSIs) with molecular profiles (mutations, copy number variation, and RNA sequencing data) across 14 cancer types and 5,720 patients. The architecture uses Attention-based Multiple Instance Learning (AMIL) for WSIs, Self-Normalizing Networks for genomics, and a Kronecker product fusion layer that captures multiplicative interactions between modalities. The multimodal model achieved a concordance index (c-Index) of 0.644 versus 0.606 for genomics-only and 0.578 for WSI-only, demonstrating that morphological and molecular information capture fundamentally different aspects of tumor biology. The improvement magnitude of approximately 6–11% mirrors the improvements observed in this dissertation when adding image data to digital sensor data. In closely related work, Chen et al. [22] used CNNs, Graph Convolutional Networks (GCNs), and gating-based attention to achieve c-Index 0.826

for glioma survival prediction, substantially outperforming individual modalities by modeling spatial relationships between histopathological features and genomic markers.

Vanguri et al. [23] fused CT radiomics, PD-L1 immunohistochemistry slides, and genomic features for prediction of immunotherapy response in non-small cell lung cancer patients. The multimodal model achieved an area under the receiver operating characteristic curve (AUC) of 0.80 versus 0.73 for PD-L1 alone and 0.61 for tumor mutational burden alone. This result has direct clinical significance because no single biomarker reliably predicts immunotherapy response, and the improvement from fusion demonstrates that each biomarker captures different aspects of the tumor–immune interaction. This finding parallels the central thesis of this dissertation: no single measurement modality can fully characterize a complex physical system, and the fusion of complementary data streams captures information inaccessible to any individual sensor.

Chen et al. [24] introduced the Multimodal Co-Attention Transformer (MCAT), using transformer cross-attention to learn dense mappings between WSI patch embeddings and genomic features. MCAT achieved superior survival prediction by explicitly modeling which histology regions are relevant given specific genomic features, representing a key architectural innovation in cross-modal attention. Hemker et al. [25] extended this concept through HEALNet, a framework for heterogeneous biomedical data fusion that scales to arbitrary numbers of modalities with instance-level interpretability. The cross-attention mechanism in MCAT has since been adopted across multiple domains, and its principle, dynamically learning which features in one modality are relevant given features in another, is conceptually related to the skip connection architecture used in this dissertation.

Huang et al. [26] conducted a systematic review of 985 studies on fusion of medical imaging with electronic health records using deep learning. Their analysis found that joint fusion, where modalities share intermediate representations, consistently outperformed early fusion (concatenation at the input level) and late fusion (combining predictions from separate models) across applications including pulmonary embolism detection, Alzheimer's prediction, and breast cancer risk assessment. This finding directly informed the hybrid parallel-series architecture used in this dissertation, which employs a joint fusion strategy where image features extracted through convolutional layers and digital/acoustic features are concatenated at an intermediate representation level before being processed through shared fully-connected layers.

Of particular relevance to the acoustic component of this dissertation, Tariq et al. [27] developed a feature-based CNN fusion framework called FDC-FS, processing multiple acoustic representations (spectrograms, Mel-frequency cepstral coefficients, and chromagrams) of lung and heart sounds. Their fusion model achieved up to 99.1% accuracy for lung sound classification and 97% for heart sound classification on augmented datasets, outperforming individual feature representations. Their approach of extracting multiple statistical and frequency-domain features from acoustic signals and processing them through dedicated neural network branches is directly analogous to the quartile-based acoustic feature extraction developed in Chapter 4. Both approaches recognize that raw acoustic waveforms contain too much redundant information for efficient neural network processing, and that statistical summarization provides an effective dimensionality reduction while preserving the essential physical information content.

The medical imaging literature thus establishes several principles that carry over to the boiling heat transfer application: morphological information (from images) and state-variable information (from genomic profiles, or from digital sensors in boiling) capture different aspects of the same system; joint intermediate fusion outperforms early or late fusion; and acoustic data can be effectively compressed into statistical feature representations without significant loss of predictive power.

2.2. Autonomous Vehicles and Robotics

Sensor fusion for autonomous driving has produced some of the most quantitatively rigorous demonstrations of multimodal advantage, driven by standardized benchmarks (nuScenes, KITTI, Waymo) and the safety-critical nature of the application. The autonomous driving domain is relevant to this dissertation because it demonstrates the principle that multimodal fusion provides its greatest advantage precisely when individual sensor modalities degrade, a principle that translates directly to the PDVR regime in boiling, where visual observation is most compromised.

Liu et al. [28] created BEVFusion, a system that projects both camera features (via Lift-Splat-Shoot with a Swin-Tiny backbone) and LiDAR features (via voxelization and a 3D backbone) into a unified bird's-eye view (BEV) representation space. On the nuScenes benchmark, BEVFusion achieved +13.6% mean Intersection over Union (mIoU) over LiDAR-only and +6% mIoU over camera-only for BEV map segmentation. The most significant results emerged under adverse conditions, where fusion achieved +10.7% NDS (nuScenes Detection Score) in rain and +12.8% NDS at night compared to single-modality approaches. These robustness improvements under degraded conditions directly parallel the finding in this dissertation that acoustic data provides its greatest predictive advantage in the PDVR regime, where visual observation is most compromised by steam obscuration and chaotic liquid dynamics.

Liang et al. [29] demonstrated that camera-LiDAR fusion improved PointPillars detection by 18.4% mAP and 10.6% NDS, and surpassed prior state-of-the-art by 15.7–28.9% mAP under simulated LiDAR malfunction. This result, that multimodal fusion provides disproportionate benefit when one sensor modality degrades, is a fundamental cross-domain principle. In boiling systems, the analogous situation occurs at high superheats, where the chaotic two-phase morphology of the PDVR regime compromises visual observation at the same time that it produces the most intense and information-rich acoustic emissions. Nabati et al. [30] further demonstrated modality-specific contributions through CenterFusion, achieving +12% NDS by adding radar to camera data, with radar's unique contribution being velocity estimation in all weather conditions. This parallels how acoustic data in boiling provides temporal dynamic information (nucleation frequency, bubble collapse intensity) that is fundamentally inaccessible to spatial imaging.

Chitta et al. [31] developed TransFuser, achieving state-of-the-art autonomous driving performance by fusing camera, LiDAR, and GPS through transformer-based multi-resolution attention mechanisms. TransFuser outperformed all prior work on the CARLA leaderboard in driving score. The key architectural insight is that cross-modal attention mechanisms can dynamically weight the contribution of each modality based on the current context, for example, allocating more weight to camera features at a traffic light (where cameras provide unique semantic information) and more weight to LiDAR features in dense fog (where cameras are unreliable). This attention-based weighting is conceptually similar to the permutation importance analysis conducted in Chapter 4, which reveals that different input features contribute differently to heat flux prediction versus regime classification.

Liu et al. [32] demonstrated through ManiWAV that vision-only robot manipulation achieved only 20% success rate on material-sensitive tasks versus significantly higher success with audio-visual fusion. The audio modality captured physical interaction dynamics, such as the difference in sound between grasping a ceramic cup and a plastic cup, that are invisible to cameras. Chen et al. [33] demonstrated a similar principle for audio-visual navigation, where agents learn to use reverberating audio to infer room geometry invisible to cameras. These results directly parallel the observation in this dissertation that acoustic emissions during boiling capture temporal dynamics of bubble nucleation and collapse at timescales faster than the camera frame rate, providing information about the boiling process that is genuinely inaccessible through imaging alone regardless of camera quality or temporal resolution.

The autonomous driving literature thus provides strong evidence for two principles central to this dissertation: multimodal fusion is most valuable precisely when individual modalities are degraded, and different modalities capture fundamentally different physical quantities (spatial geometry versus semantics in driving; spatial morphology versus temporal dynamics in boiling) that provide complementary rather than redundant information.

2.3. Industrial Process Monitoring and Manufacturing

The industrial process monitoring domain provides the most direct technological parallel to the boiling heat transfer application investigated in this dissertation. Additive manufacturing, welding, and machining processes share key characteristics with boiling: complex and stochastic physics governing the process; multi-sensor instrumentation combining cameras, acoustic sensors, and temperature measurements; and the need for real-time quality prediction from fused sensor data. The architectural choices made in this domain are therefore the most directly transferable to the boiling monitoring framework.

Petrich et al. [34] achieved 98.5% accuracy in classifying additive manufacturing build quality by fusing layerwise optical imagery, acoustic emissions, multi-spectral photodiode data, and scan vector trajectory data from laser powder bed fusion. Their architecture used separate feature extraction branches for each modality followed by concatenation and classification, achieving a 4.2% improvement over the best single-modality result. Sensitivity analysis revealed that while optical imagery contained the highest individual information content, additional modalities significantly improved performance beyond any single modality. The acoustic emissions in their system captured solidification dynamics, crack formation, gas release, and powder ejection, that are invisible to surface-only optical sensors, directly analogous to how acoustic emissions in boiling capture subsurface bubble collapse dynamics invisible to cameras. In earlier foundational work, Shevchik et al. [35] pioneered acoustic-only monitoring for additive manufacturing with spectral CNNs, achieving 83–89% accuracy across three porosity levels and establishing that acoustic emissions provide depth-penetrating information inaccessible to optical methods.

Xia et al. [36] achieved 99.75% fault detection accuracy for machine tool spindle bearings by fusing vibration and acoustic emission data through parallel one-dimensional CNN branches. Their approach of using separate CNN branches for different one-dimensional signal types, followed by feature concatenation and joint classification, provides direct architectural precedent for the acoustic processing branch in the multimodal fusion network developed in Chapter 4.

Jiang et al. [37] developed SIMNet for weld penetration monitoring, fusing arc sound with weld pool images using cross-attention mechanisms and achieving penetration width prediction with mean squared error of 0.1141 mm at 60 frames per second. The cross-attention approach used in SIMNet, which learns alignments between audio features and visual features across time, represents an advanced fusion strategy that could be adapted for boiling analysis, where the temporal alignment between acoustic events (bubble nucleation, collapse) and visual changes (bubble appearance, dryout formation) carries physical significance. Kullu et al. [38] further established the principle of domain-dependent modality importance by demonstrating that different fault types in rotating equipment are best detected by different signal domains, inner ring faults in current spectrograms, outer ring faults in vibration spectrograms, paralleling the finding in Chapter 4 that different boiling regimes benefit differentially from acoustic versus visual information.

The industrial process monitoring literature establishes the direct precedent for the multimodal boiling monitoring architecture developed in this dissertation: parallel sensing branches for image, acoustic, and thermal data; intermediate fusion through concatenation; and the principle that acoustic emissions capture subsurface physical phenomena that are invisible to optical sensors. The time-frequency transformation of one-dimensional signals (acoustic, vibration) into two-dimensional spectrograms for CNN processing, demonstrated across multiple manufacturing papers, also provides an alternative to the quartile-based feature extraction developed in Chapter 4, one that preserves more spectral detail at the cost of higher dimensionality.

2.4. Environmental and Geophysical Monitoring

Environmental monitoring applications demonstrate how multimodal fusion handles heterogeneous data from diverse sensor platforms operating at fundamentally different spatial and temporal scales. The domain is relevant to this dissertation because it establishes the principle that the fusion of modalities capturing different physical quantities (spectral signatures versus geometric structure; seismic velocity versus geological features) consistently outperforms single-modality approaches, even when the modalities have vastly different dimensionalities and sampling characteristics.

Li et al. [39] provided a comprehensive 46-page review documenting that deep learning in multimodal remote sensing data fusion spans spatiospectral, spatiotemporal, LiDAR-optical, and SAR-optical fusion approaches. Their analysis established that hyperspectral-LiDAR classification accuracy improved from 80.39% (hyperspectral only) to 89.60% (fused), a 9.21 percentage point improvement achieved through dual-branch networks with cross-attention modules. The hyperspectral data captures spectral signatures (chemical composition, vegetation health) while LiDAR captures three-dimensional geometric structure (canopy height, terrain elevation). Neither modality alone can distinguish land cover types that differ in only one characteristic, for example, two vegetation types at the same height with different spectral signatures, or two materials at different heights with similar spectra. The fusion captures both dimensions simultaneously, analogous to how visual and acoustic data in boiling capture spatial morphology and temporal dynamics respectively. A 2025 study further demonstrated that triple fusion (optical + SAR + LiDAR) achieved 83% accuracy versus approximately 71% for any two-sensor combination, establishing diminishing but still meaningful returns from adding additional modalities, consistent with the progressive improvement observed in this dissertation from digital-only to image-plus-digital to image-plus-digital-plus-acoustic.

Mousavi et al. [40] introduced EQTransformer, an encoder-decoder architecture with hierarchical attention combining global transformer attention for earthquake detection and local attention for seismic phase picking. Trained on 1.3 million seismic waveforms, EQTransformer detected twice as many earthquakes as the existing catalog using only one-third of the seismic stations. The transformer architecture's ability to capture long-range temporal dependencies in waveform data is particularly relevant for acoustic boiling analysis, where the temporal pattern of acoustic events, not just their individual characteristics, carries information about the boiling regime and its evolution over the course of a vaporization event. In related work, Hu et al. [41] developed SafeNet for earthquake forecasting by fusing 282-dimensional seismic indicators with geological/tectonic maps and spatial earthquake distributions, outperforming 13 state-of-the-art models and demonstrating successful cross-region transfer from China to the United States, establishing the generalization potential of multimodal geophysical models.

Yuan et al. [42] comprehensively reviewed deep learning improvements across atmospheric, vegetation, hydrological, and ocean monitoring applications, documenting reduced uncertainty compared to physical-model-based approaches. Their finding that data-driven approaches can complement and improve upon physics-based models is particularly relevant for boiling heat transfer, where physics-based correlations have served as the primary analytical tool for decades but face limitations in capturing the complex, regime-dependent behavior observed in nanostructured surface vaporization experiments. A 2025 study in forest and environmental monitoring demonstrated that fusion of LiDAR three-dimensional structure with optical spectral features improved soil moisture estimation to $R^2 = 0.891$ and canopy height estimation to $R^2 = 0.98$, further establishing that complementary physical measurements (structure and spectrum, analogous to morphology and acoustic dynamics in boiling) provide additional information.

The environmental monitoring domain contributes the following principles to the boiling heat transfer analysis: first, modalities capturing different physical quantities provide complementary information even when their sampling characteristics differ dramatically; second, attention-based architectures can effectively handle the temporal alignment challenges that arise when fusing data streams sampled at different rates; and third, the progressive addition of sensing modalities yields diminishing but still meaningful returns.

2.5. Speech and Audio-Visual Processing

The speech and audio-visual domain provides foundational work on fusing audio with visual data, the same fundamental modality combination underlying the acoustic-visual fusion developed in this dissertation for boiling analysis. This domain has pioneered many of the architectural innovations (cross-modal attention, self-supervised pre-training, contrastive learning across modalities) that have subsequently been adopted across other application domains.

Shi et al. [43] introduced AV-HuBERT, a hybrid ResNet-Transformer framework that learns shared audio-visual representations through self-supervised masked multimodal cluster prediction on lip video and audio. With only 30 hours of labeled data, AV-HuBERT achieved 32.5% word error rate (WER) on the LRS3 benchmark, outperforming prior state-of-the-art (33.6% WER) that required 31,000 hours of labeled data, a 1,000-fold data efficiency improvement through self-supervised multimodal learning. Under noisy conditions (0 dB signal-to-noise ratio), the audio-visual model reduced WER by 50% compared to audio-only models. This dramatic improvement under degraded conditions directly parallels the finding in this dissertation that acoustic data provides its greatest benefit in the PDVR regime, where visual observation is most compromised by steam obscuration and chaotic liquid dynamics.

Girdhar et al. [44] developed ImageBind, which binds six modalities (images, text, audio, depth, thermal, and IMU data) through a single shared embedding space using image-paired contrastive learning. The key innovation is that explicit pairing between all modality combinations is not required during training, only image-paired data for each modality is needed, because images serve as a universal binding modality. ImageBind achieved emergent zero-shot classification of 77.7% on ImageNet, 66.9% on ESC audio, and 54.0% on NYU-D depth. This approach could potentially be adapted for boiling heat transfer, where paired image-acoustic-thermal data could enable training of unified representations that capture the relationship between visual morphology, acoustic signatures, and thermal state without requiring all three modalities to be simultaneously available for every training sample.

Tsai et al. [45] introduced the Multimodal Transformer (MulT) for unaligned multimodal language sequences, using directional pairwise cross-modal attention across text, visual, and acoustic modalities. Six cross-modal transformers model all modality pairs, handling different sampling rates without explicit alignment. MulT outperformed state-of-the-art by 5–15% across three benchmark datasets. The explicit handling of different sampling rates is a critical consideration for the boiling application in this dissertation, where acoustic data is sampled at 48 kHz while video frames are captured at 1,200 fps, representing fundamentally different temporal resolutions. The cross-modal attention mechanism in MulT learns which time steps in one modality are most relevant to which time steps in another, providing a principled approach to temporal alignment that could enhance future versions of the boiling multimodal architecture.

Lian et al. [46] surveyed deep learning-based multimodal emotion recognition across facial expressions, speech audio, and text, finding that trimodal systems achieve 70–80% weighted accuracy versus 50–65% for unimodal baselines. Liu et al. [47] achieved 76.76% accuracy on the RAVDESS benchmark using adaptive cross-modal attention, a +20.2% improvement over audio-only, by selectively reinforcing each modality with only complementary (not redundant) information from other modalities. This selective reinforcement principle is relevant to the boiling application because the image and acoustic modalities in boiling capture partially overlapping information (both reflect overall nucleation intensity) as well as complementary information (spatial morphology versus temporal dynamics), and an optimal fusion strategy should emphasize the complementary information while avoiding amplification of shared noise.

The audio-visual domain thus contributes foundational architectural innovations, cross-modal attention for handling different temporal resolutions, self-supervised pre-training for data-efficient learning, and contrastive learning for binding heterogeneous modalities, that represent the frontier of multimodal learning and provide a roadmap for future improvements to the boiling heat transfer framework.

2.6. Structural Health Monitoring

Structural health monitoring (SHM) shares key characteristics with boiling process monitoring: both involve non-destructive sensing of dynamic physical systems, both face the challenge of extracting meaningful features from noisy sensor data, and both benefit from combining sensors that capture complementary physical phenomena (vibration and acoustic emission in SHM; visual morphology and acoustic emission in boiling). The SHM domain is particularly relevant because it provides extensive precedent for acoustic emission analysis using deep learning architectures.

Azimi et al. [48] established through a comprehensive review that CNNs demonstrate superior noise resistance compared to SVM and KNN for structural damage classification, while Jia et al. [49] systematically reviewed 337 articles published between 2017 and 2023 on deep learning for SHM, finding that CNNs dominate the SHM literature (approximately 60% of studies), vibration and image data account for 80% of studies, and crack detection is the most studied problem. Hassani et al. [50] documented through a systematic review of data fusion techniques for SHM that CNN-based fusion recovered missing strain data with greater than 80% accuracy even when only one of four sensors remained functioning, demonstrating the fault-tolerance capability of multimodal systems.

Yan et al. [51] introduced a Quadratic Convolution Neural Network (QCNN) for audio-vibration fusion in bearing fault diagnosis, achieving up to 99.99% sample-level accuracy for single-bearing fault classification across different working speeds, and exceeding 98.81% for multiple fault types even under 0 dB noise conditions. The quadratic neurons inherently model nonlinear interactions between input features, suggesting that multiplicative fusion strategies may capture interaction effects missed by simple concatenation.

The SHM domain thus provides direct architectural precedent for the acoustic processing components of the multimodal boiling analysis: parallel one-dimensional CNN branches for acoustic signals, feature-level concatenation for multi-sensor fusion, and the principle that combining spatially-resolved and temporally-resolved sensing modalities captures complementary damage (or boiling) phenomena.

2.7. Cross-Domain Principles

The review of prior work across six diverse technology domains reveals several consistent principles that directly informed the multimodal architecture developed in this dissertation. These principles emerge independently across all six domains, suggesting that they reflect fundamental properties of multimodal information rather than domain-specific artifacts.

The first principle is that complementary information is the core driver of multimodal advantage. In every domain reviewed, the improvement from fusion arises because different sensor modalities capture fundamentally different physical phenomena: morphology versus molecular state in medicine; geometry versus semantics in autonomous driving; surface dynamics versus subsurface events in manufacturing; spectral signatures versus geometric structure in environmental monitoring; lip movements versus acoustic speech signals in audio-visual processing; vibration versus acoustic emission in structural health monitoring. In boiling heat transfer, visual data captures the spatial morphology of the bubble–liquid–surface system, acoustic data captures the temporal dynamics of nucleation, growth, coalescence, and collapse events, and digital sensors capture the boundary conditions (temperature, droplet volume, contact angle) that constrain the physical system. The consistent 5–20% improvement across domains when fusing complementary modalities provides the theoretical basis for the progressive improvement from 15.8% to 4.6% RMSPE observed in this dissertation.

The second principle is that intermediate fusion consistently outperforms both early and late fusion. This finding was explicitly documented in the medical imaging systematic review [26], and validated in the autonomous driving literature [28,29]. Early fusion (concatenating raw inputs) forces the network to simultaneously learn modality-specific features and cross-modal relationships, which is inefficient when modalities have vastly different dimensionalities (as with 100×100 pixel images versus 12 acoustic features in this dissertation). Late fusion (combining predictions from separate models) prevents the network from learning cross-modal interactions that may be essential for accurate prediction. The skip connection architecture employed in this dissertation represents an intermediate fusion strategy where each modality is first processed by modality-specific layers (convolutional layers for images, the acoustic branch for quartile features) before being concatenated at a shared fully-connected layer that can learn cross-modal relationships.

The third principle is that robustness under degraded conditions is where multimodal fusion provides its greatest advantage. BEVFusion's 15.7–28.9% mAP improvement under LiDAR malfunction [29], AV-HuBERT's 50% word error rate reduction under noise [43], and the SHM studies showing greater than 80% accuracy recovery with sensor loss all demonstrate that fusion's value peaks when individual sensors fail or degrade. For boiling systems, this principle is critical because the PDVR regime, the most challenging and potentially dangerous regime for thermal management, is precisely where visual observation degrades most due to steam obscuration and chaotic liquid dynamics, and where acoustic data provides the most complementary information about vapor recoil events and intermittent surface rewetting.

The fourth principle is that time-frequency transformation bridges heterogeneous data types. Converting one-dimensional signals (acoustic, vibration) to two-dimensional representations via short-time Fourier transform (STFT) or wavelet transforms enables unified CNN processing with visual data, as demonstrated across speech [43], and geophysics [40] applications. The quartile-based feature extraction developed in Chapter 4 represents an alternative dimensionality reduction strategy that preserves the essential statistical characteristics of the acoustic signal while dramatically reducing data volume. The relative merits of spectrogram-based versus statistical feature extraction depend on the specific application: spectrograms preserve fine-grained frequency content but require higher-dimensional processing, while quartile features are more computationally efficient and provide physically interpretable summary statistics.

The fifth principle is that physics-informed hybrid approaches outperform purely data-driven methods for extrapolation to unseen conditions. This has been documented across nuclear thermal-hydraulics, condensation monitoring, and industrial process control. Physics-informed neural networks (PINNs) and physics-guided architectures embed conservation laws, constitutive relationships, or physical constraints directly into the loss function or network architecture, improving both interpolation accuracy and generalization capability. The architecture developed in this dissertation incorporates physics-informed elements implicitly rather than explicitly: the choice of input features (superheat temperature, droplet volume, contact angle) reflects the key physical parameters governing heat transfer performance; the regime classification output incorporates the physical understanding that boiling progresses through distinct regimes; the heat flux calculation embeds energy conservation; and the skip connection architecture reflects the physical understanding that boundary condition data and morphological data provide complementary information. Future work could more explicitly incorporate physics constraints, as will be discussed in Chapter 6.

These five cross-domain principles, complementary information, intermediate fusion, degraded-condition robustness, time-frequency bridging, and physics-informed learning, provide the theoretical foundation for the multimodal boiling heat transfer analysis developed in Chapters 3 and 4, and inform the assessment of enhancement in Chapter 5 and future directions in Chapter 6. The consistency of these principles across such diverse application domains provides strong evidence that the multimodal approach developed in this dissertation is not merely a technique tuned to a specific experimental configuration, but rather an instantiation of fundamental information-theoretic principles about how complementary sensor modalities should be combined for optimal prediction of complex physical systems.

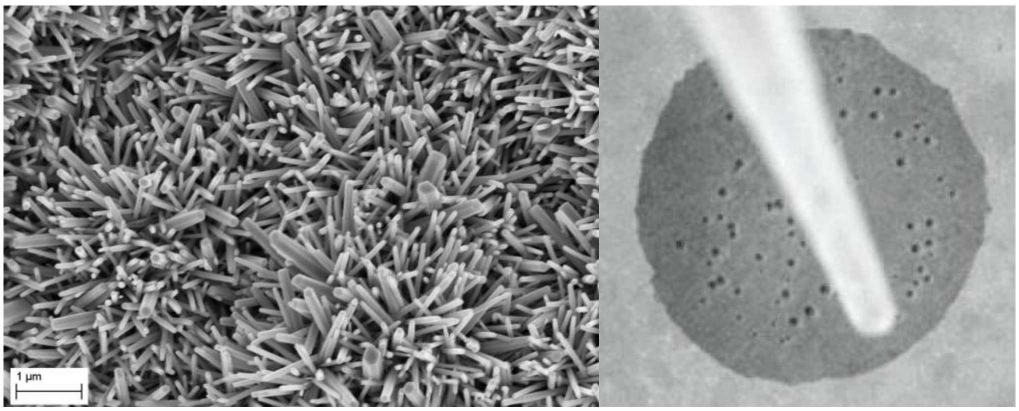
Chapter 3.
CNN-Based Combined Image and Sensor Data Analysis and Performance Modeling

3.1. Motivation: Single versus Multiple Mode Analysis and Modeling

The majority of prior machine learning work in boiling heat transfer, as reviewed in Section 1.4, has relied on a single data modality, typically digital sensor measurements such as temperature, pressure, and flow rate. While these approaches have achieved useful prediction accuracy for heat transfer coefficients and boiling curves, they fundamentally cannot capture the rich morphological information visible in the two-phase mixture on heated surfaces. The visual appearance of the boiling process, the size, shape, number, and spatial distribution of bubbles; the extent of wetted and dry regions; the dynamics of liquid film evolution, contains substantial quantitative information about the active heat transfer mechanisms that is complementary to what can be measured by traditional digital sensors. It is noteworthy, however, that until very recently, the correlation of image information with expected morphology regimes and digital measured parameter trends was usually established through manual comparisons of image features (as depicted in Figure 3.3). This approach established the now well-known combinations of boiling liquid-vapor morphology features that correspond to the isolated bubble regime, the regime of slugs and bubbles (incipient CHF regime), the transition boiling regime, and the film boiling regime acknowledged in review articles and textbooks on phase change (see, for example, [52, 53]).

The recent availability of convolution neural networks (CNN's) to analyze information content in images has opened the door to using CNN's to analyze images captured during boiling processes. Developed in the 1980s [54,55], convolution neural networks are now commonly used as the backbone of many computer vision applications. These applications range from self-driving vehicles to medical image analysis, facial recognition, and more. Recent studies of boiling heat transfer have explored ways to leverage this powerful tool to better understand boiling processes. The paper of Yang, et al. [15] discusses the use of high-speed camera visualization to examine flow boiling in microchannels. The study aims to use a CNN model to identify the flow pattern for use as a means of accurate heat transfer mechanism prediction for applications. To achieve an optimized performance, the image was pre-processed to address issues related to unbalanced lighting and other effects that may not be relevant to the classification of pattern flow. In another similar study, Lee [16] explored strategies to optimize experimental observation and measurements of pool boiling heat transfer using computer vision techniques. The computer vision applied in that study performs bubble detection and segmentation to track bubble movements as they depart from the surface. This helps in capturing key performance parameters such as bubble density, size, and departure frequency autonomously.

While recent studies in this area have considered collecting digital data from sensors and observation or imaging of the two-phase morphology, CNN analysis of boiling process images has often been done in parallel with an analysis to fit a physics-based model, or a machine learning model (such as a neural network), to the digital data from the experiment. Because past research has established a correlation between superheat level and boiling regimes, separate CNN model analysis of boiling regimes and two-phase morphology has frequently postulated them to be functions primarily of superheat. If this postulated connection is adopted, training of a CNN model that predicts superheat from image features is straightforward and valuable in cases where the image features are strongly correlated to superheat alone, the CNN model can achieve a good fit to the data. The research presented in this chapter addresses this limitation by developing CNN architectures designed to simultaneously analyze both digital sensor data and high-speed video images captured during droplet vaporization experiments on ZnO nanostructured surfaces.



(a) (b)

Figure 3.1. (a) A micrograph of the nanostructured surface. (b) Image of a vaporizing droplet.

A micrograph of the nanostructured surface type considered here is shown in

Figure 3.1. Earlier investigations of pool boiling and droplet vaporization on surfaces of this type have indicated that droplet vaporization on these surfaces exhibits distinct regimes of heat transfer, similar to those for pool boiling. Figure 3.2. illustrates the correspondence between droplet vaporization regimes and pool boiling regimes documented in the experimental studies summarized by Wemp, et al. [1, 56] and Carey, et al. [4]. Vaporization of water droplets deposited on the nanoporous surface shown in

Figure 3.1 is of interest because the surface is hydrophilic, exhibiting a contact angle less than 10 degrees, and the surface exhibits extreme wicking. At low to moderate surface superheat levels, this causes impinging droplets to spread rapidly after first contact to form a very thin liquid film with a large footprint. The subsequent vaporization of the liquid film occurs rapidly, resulting in a high mean heat flux during the vaporization process. The schematic depiction of this process is shown in Figure 3.2. The boiling curve depicted in this figure shows the benefit of using a super-hydrophilic nanostructured surface, as it leads to increased critical heat flux and a higher corresponding superheat value.

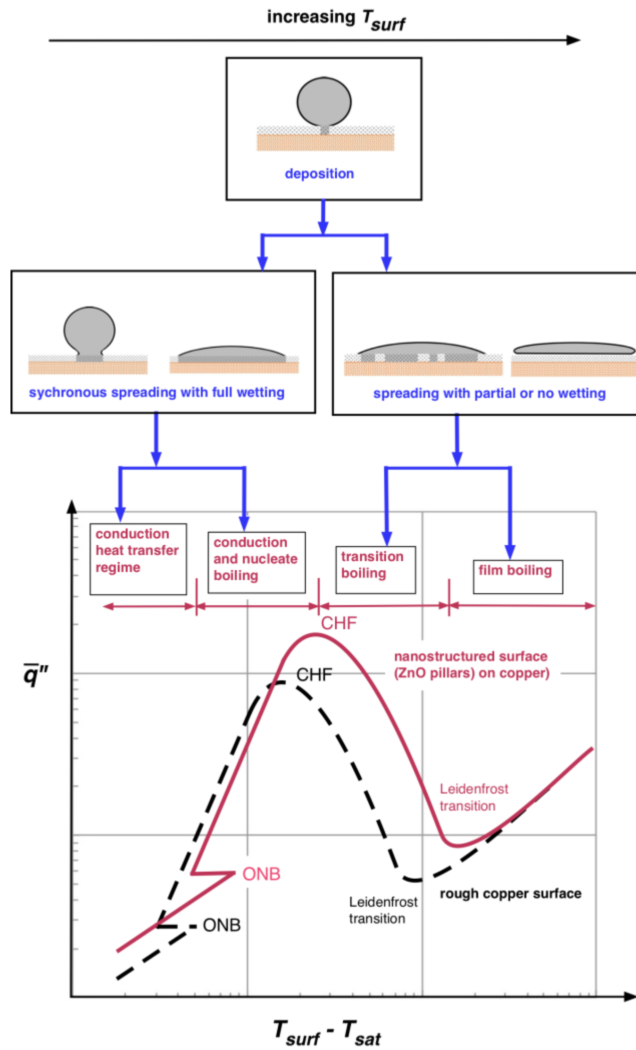


Figure 3.2. Droplet vaporization regimes and pool boiling regimes correspondence.

The work discussed in this chapter has been published in the ASME Heat Transfer Summer Conference [18] and the ASME Journal of Heat and Mass Transfer [5]. The fundamental hypothesis is that a CNN model trained on combined digital and image data can achieve more accurate predictions of heat transfer performance than a model trained on digital data alone, because the image data encodes morphological information about the boiling process that is not fully captured by the digital measurements. This type of deposited droplet vaporization is particularly well-suited for exploring the potential advantages of simultaneous machine learning analysis of digital and image data for two reasons. First, this process has wider variations of two phase morphology than the nucleate pool boiling processes considered by Suh, et al. [20], Hobold and da Silva [33], and Wu, et al. [23] which focused on bubble ebullition associated with nucleate boiling over low to moderate heat flux levels. Prior investigations have demonstrated that for the droplet vaporization process, the system exhibits several distinctly different two-phase morphologies as surface superheat varies, which makes it a more challenging task for the machine learning model. The second advantage is that for this type of experiment, optical access is easily achieved by mounting the camera to take high-speed videos from a viewpoint directly above the heated surface.

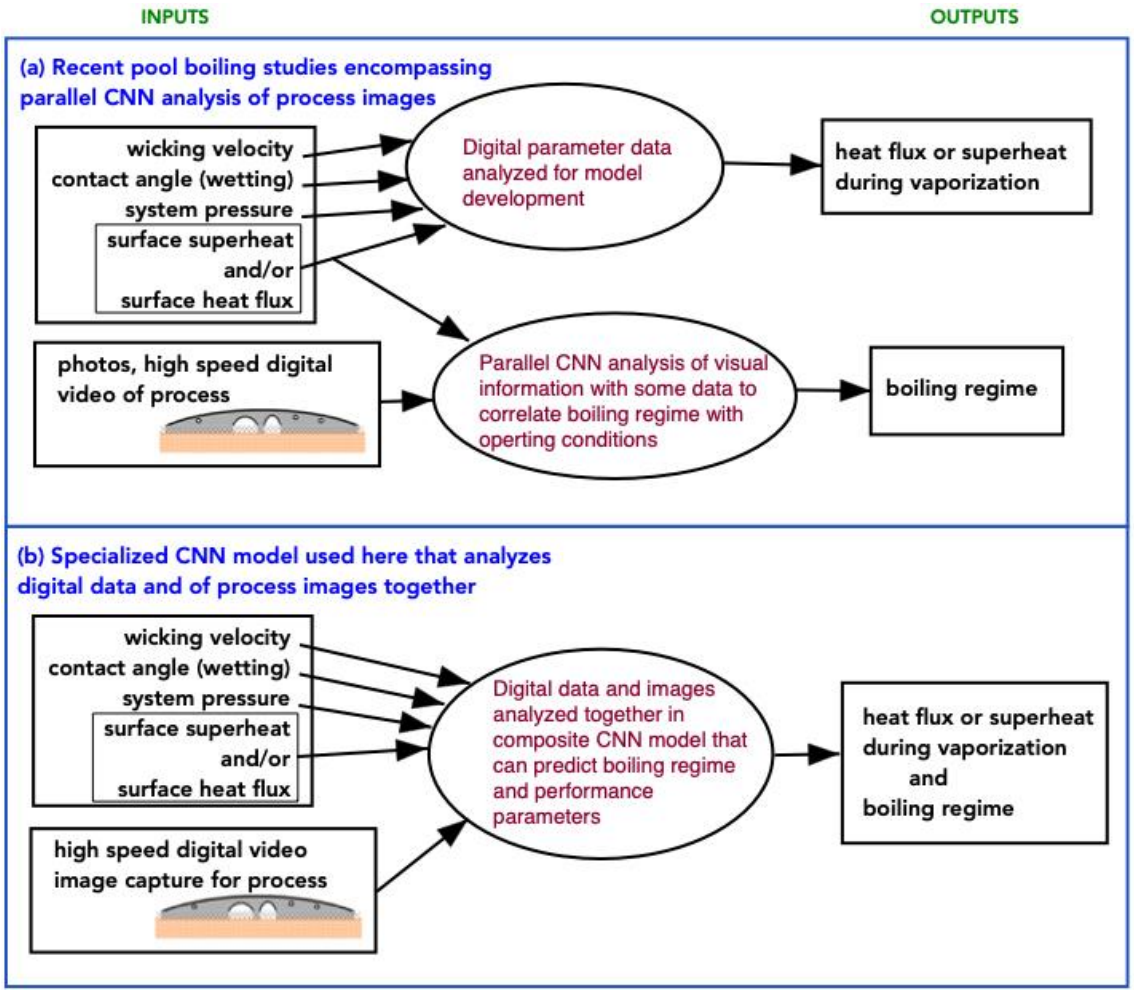


Figure 3.3. Categories in the evolution of data analysis and morphology assessment in boiling

The three architectures developed, Cases A, B, and C, progressively explore the value of incorporating different data modalities, culminating in a hybrid parallel-series CNN design with skip connections that achieves 5.8% RMSPE on testing data, representing a 57.4% improvement over the digital-only baseline.

3.2. Simulated Data Exploration

Before applying CNN models to real experimental data, the feasibility of the combined image-and-digital-data approach using a database was created with 138 images and their corresponding operating conditions and parameter values. The trends in this database matched the trends observed in the data from the earlier study by Carey, et al. [5]. The simulated images were created using a drawing tool called EazyDraw. The simulation exhibits specific features we may observe in a real physical system such as a change in morphology with respect to a change in its physical condition. Figure 3.4 depicts the different morphology observed, which are spherical cap, thin continuous line, irregular thin continuous line, and shattered regime. As shown in Figure 3.3, the CNN model task is to capture the morphology features in the image data as well as the physical conditions of the system to make predictions of the mean heat transfer rate. The physical condition of the system explored has a wall superheat temperature that ranges from 0.87 K to 56.4 K above saturation temperature and a relative atmospheric pressure that ranges from 0.47 to 2.13. Both the superheat temperature and atmospheric pressures are needed to adequately make predictions of the mean heat transfer rate.

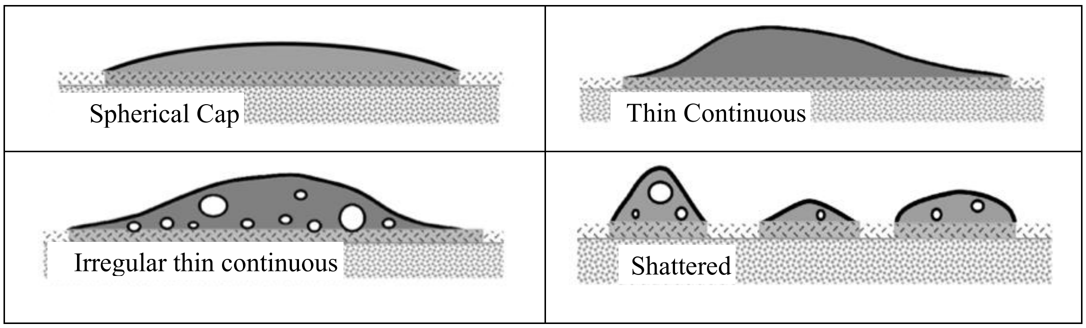


Figure 3.4. Simulated Image Morphology Classification

In the first step of this investigation, the image data and relative atmospheric pressure were used as input data for the CNN and the wall superheat temperature was used as output data. To train the CNN model efficiently, the data needed to be preprocessed. Python, with its useful open-source libraries, was the programming language used in this study. The input and output data were preprocessed to ensure the model interacted with data of similar ranges of values. The original images were converted to grayscale since the coloration was consistent for all images. This operation also helps to minimize the computational cost since greyscale images are represented by only one layer of numerical values as compared to color scale, which is represented by 3 layers. The image input column was converted to an image array of numerical values using the cv2 package, a package widely used for computer vision applications in Python. The images were resized to 200x50 pixels. The value of the image input array was divided by 255 since pixels in computers are encoded in values ranging from 0 – 255. This results in an input image array with values that lie between 0 and 1. The other input column consists of the relative atmospheric pressure whose median value was 1 and did not require preprocessing. On the other hand, each image had an associated morphology class label and wall superheat temperature. These were used as output values to train the CNN. The morphology class labels were converted to numerical values using one_hot_encoding from Keras.utils library. The superheat values were standardized by dividing them by the median superheat. In the case of this investigation, the median superheat value was 19.8 K. By performing these preprocess transformations, most of the data values were around the vicinity of 1. The entire dataset was arranged in a data frame using the Pandas package, widely used to manipulate tabular data.

Since the amount of input data was very low, there was a need to adopt a strategy to remedy this condition. Performing data augmentation is a widely adopted strategy for this kind of situation. The process consists of performing several small modifications to each image in the dataset such as a horizontal shift of the image to the right or left. This operation helps to increase the robustness of the CNN model. ImageDataGenerator from Keras was used in this study. The parameter of the generator includes random rotation of up to 5 degrees. It also includes horizontal shift, vertical shift, shear, and zoom of up to 10%. Random horizontal flips were allowed to be performed on the images as well. Empty spaces due to the image modification were filled using the nearest pixel value strategy. The resulting data augmented images were added to the original 138 images totaling 3588 image data. These images with their corresponding labels were ready to be trained. The model architecture to train this data needed to be customized in such a way that it could simultaneously take the image and relative atmospheric pressure data as input. Also, the model needs to simultaneously predict the morphology class and the wall superheat temperature.

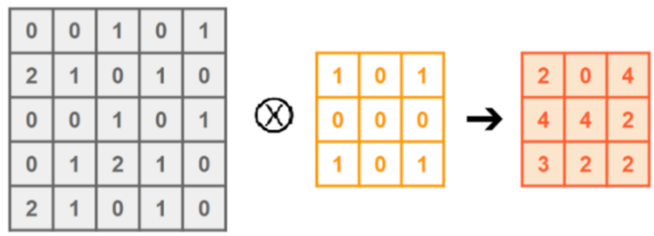


Figure 3.5. Convolution operation example

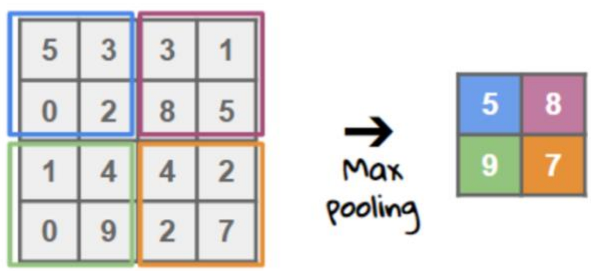


Figure 3.6. Max pooling example

For higher flexibility in building the desired convolution neural network model, the Keras functional API (Application Program Interface) was used. This API allows each layer of the network to be initialized as a function. This allows for a flexible implementation of increased complexity in the model development. The model developed here had to consider that 1 of the inputs is an array of numbers with the shape 200x50 representing the image input while the other input is a variable representing the relative atmospheric pressure. There are 2 outputs to be predicted by the model. One of the outputs requires a classification approach while the other needs a continuous variable approach such as regression. The resulting model architecture to fit this requirement is shown in Figure 3.7. This architecture is divided into four regions. The first region is the input region. This is where the image and relative atmospheric pressure are captured. The second region is the convolution region. This region processes the image input and captures the features by passing the 3x3 convolution window with 32 filters across the image input with a stride of 1, or one step at a time, and a same padding type, meaning that a padding is added to the input image to ensure that the output, after the convolution process, have the same shape as the input image. A convolution layer attempts to capture patterns such as the edges of an object present in the image by passing a convolution window onto the image array one step at a time. An example of this operation is shown in Figure 3.5. To add nonlinearity to the process, a Rectified

Linear Unit (ReLU) activation function is applied to the output of the convolution operation. A 2x2 Max pooling layer of stride 2 and a same padding type is applied to summarize the features captured. An example of this operation is shown in Figure 3.6. A same padding in max pooling allows padding to be added as needed to make up for an imperfect fit for the max pooling operation. This set of three operations is repeated once to capture more complex features of the image.

The output from the two stacked sets of operations is flattened upon moving to the fully connected region of the architecture. In this region, the flattened layer passes through a dense layer with 32 neurons. A ReLU activation function is applied to the layer to add non-linearity. The relative atmospheric pressure is then appended to the resulting output through a concatenation layer. From this, the data is passed through a dense layer with 4 neurons and a softmax activation function to generate the morphology classification prediction. Since we are dealing with classification and regression, the regression part of the architecture has been laid out similarly. The input image is passed to a two-stacked set of convolution operations, then gets flattened, and it is passed to a dense layer with 32 neurons and ReLU activation function. From this point, the operations differ from the classification route. The classification output and the relative atmospheric pressure inputs are appended to the result after the dense layer and ReLU activation via a concatenation layer. The concatenated layer output is passed to another dense layer with 32 neurons and a ReLU activation function. The result here is passed to the output region with one neuron and a linear activation function to predict the wall superheat temperature. With this structure, the model was trained on the 3588 images with their labels.

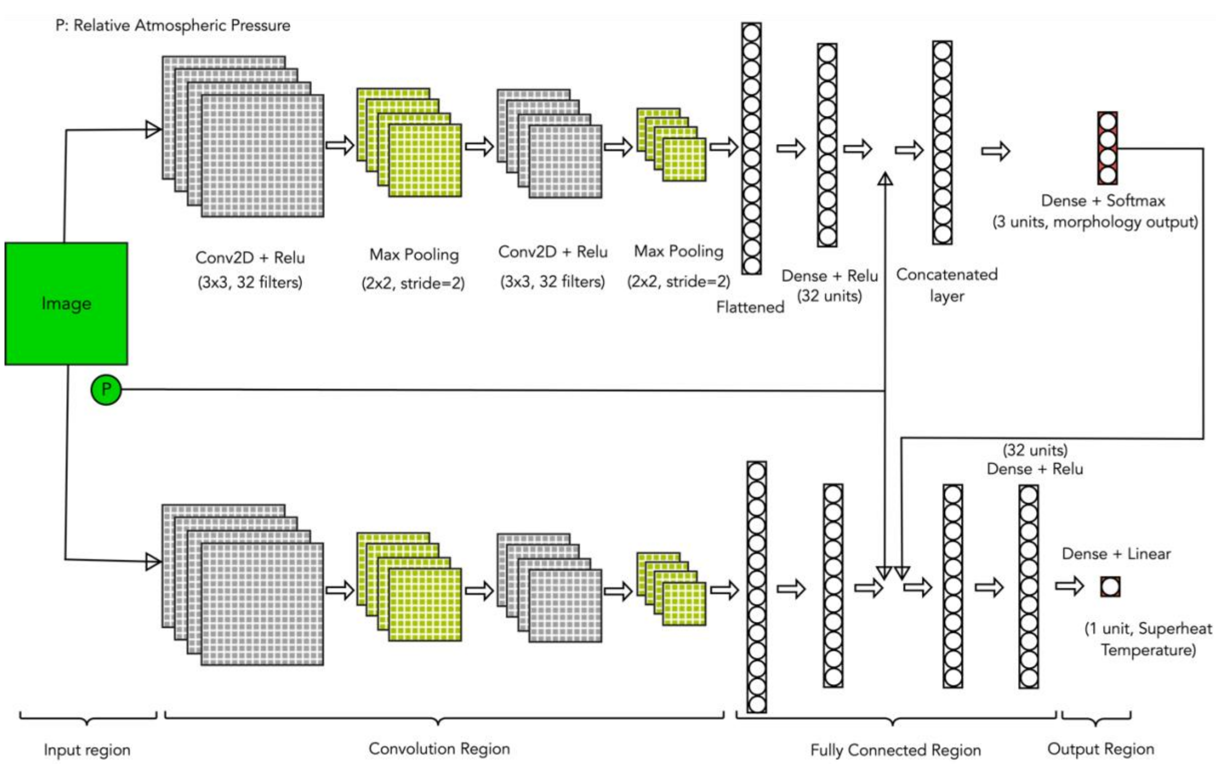


Figure 3.7. Convolution Neural Network Architecture for simulated data

The model depicted in Figure 3.7 fulfills all the requirements necessary to train the convolutional neural network to predict the morphology class and superheat temperature. An Adam optimizer was used in the model training. Adam stands for adaptive moment estimator. This optimizer helps stochastic batch processes to converge faster to an optimal solution. Its input value in this study was decreased, going from 10^{-3} to 10^{-6}, as needed to optimize the loss in the model prediction. In our case, the loss strategy used for the morphology classification was the categorical cross entropy while that for superheat prediction was the mean absolute error. Each of these losses was weighted equally in the optimization model. The training batch size was set at 32. 25% of the data was used as a validation set. The training was performed on 75% of the data. The model converged to an average loss of 0.05 on the validation set. The model was then tested on a new set of data with 100 observations generated from the simulated droplet model. Figures 6 and 7 illustrate the result obtained from the optimized convolutional neural network tested on a new set of unseen data. On the testing dataset in the simulated sample images, we observed a perfect morphology classification in all four classes of image data. Also, the wall superheat temperature prediction performed well with a root mean square percent error (RMSPE) of only 12% on a set of unseen data. This strong predictive capability suggests that the approach used in this investigation to train the model is a promising analysis of boiling process images. This is an indication that the model has been able to successfully learn relevant patterns present in the image to make accurate classifications.

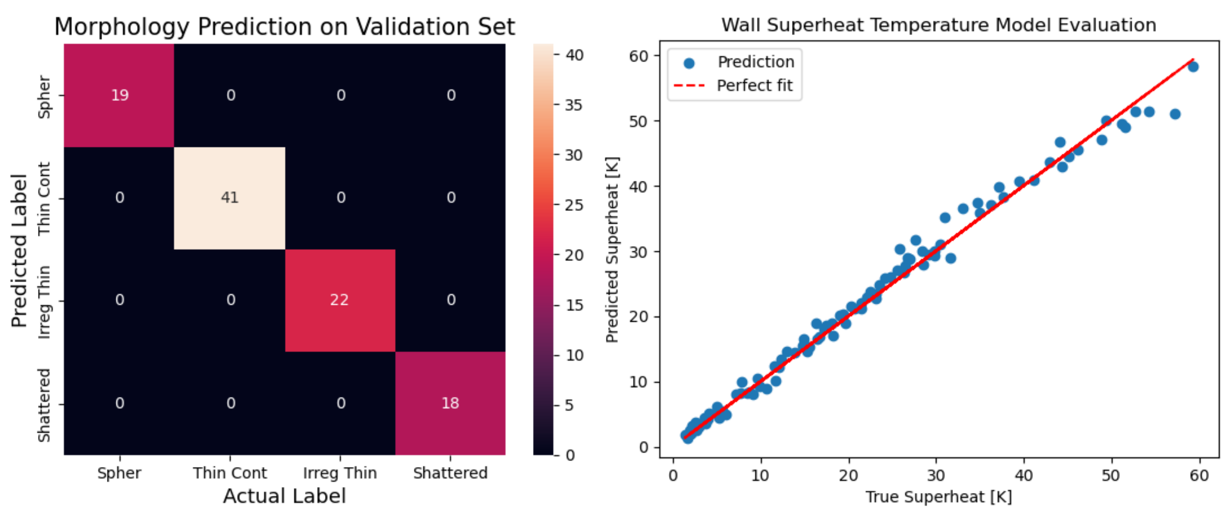


Figure 3.8. Prediction Result of Simulated Data Validating CNN as a tool for Boiling Heat Transfer Processes

3.3. Experimental Data Collection

The experimental apparatus, depicted schematically in Figure 3.9, consists of a 2.5 cm diameter cylindrical copper substrate with a ZnO nanostructured coating on its top surface, placed on a temperature-controlled hot plate with approximately 1 cm thick mineral wool insulation on the sides to ensure the copper cylinder equilibrates uniformly to the hot plate temperature. The fabrication process for the ZnO nanostructured surface follows the procedure described in Section 1.2. The resulting test surface exhibits a contact angle of approximately 6° and extreme wicking behavior.

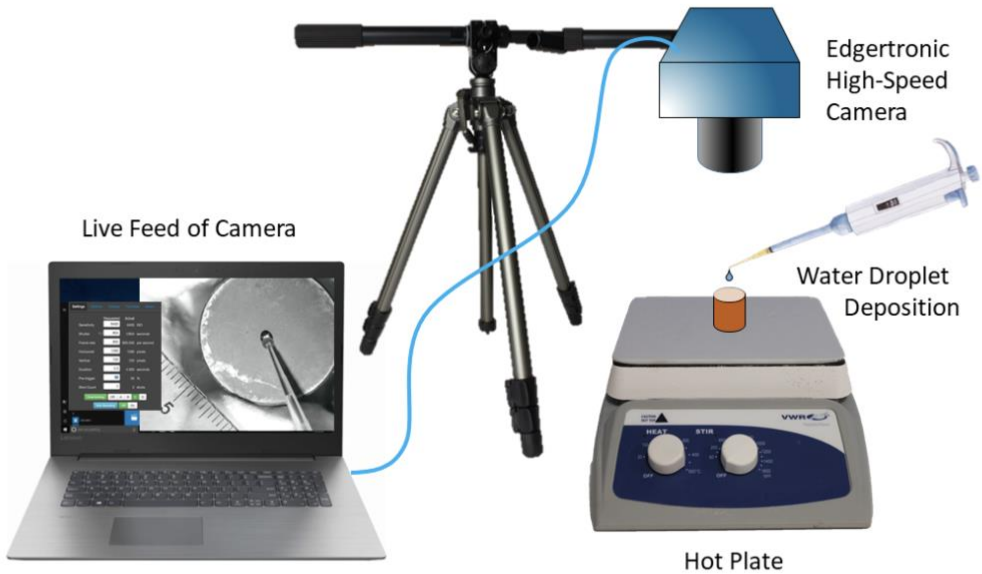


Figure 3.9. Schematic of droplet deposition/vaporization experiment.

The droplet deposition experiment employed in this study differs fundamentally from conventional pool boiling experiments in several ways that make it uniquely suited for machine learning analysis. In pool boiling, a heated surface is submerged in a large liquid reservoir, and the boiling process reaches a quasi-steady state at each superheat level, with a statistically stationary distribution of nucleation events. In contrast, the deposited droplet experiment captures a transient vaporization event from initial contact through complete evaporation, producing a time-resolved sequence of morphological states within each individual test.

When a water droplet is deposited onto the superheated ZnO nanostructured surface, the ultra-low contact angle (approximately 6°) and extreme capillary pressure within the nanopillar array cause the liquid to spread rapidly in two distinct phases. In the first phase, synchronous spreading, the liquid front within the nanostructure and the upper droplet contact line advance together, fully spreading an 8 μL droplet to a mean thickness of approximately 70 μm within the first 50 ms after deposition. In the second phase, hemi-spreading, the upper droplet settles at its equilibrium contact angle while liquid continues to wick outward through the nanostructure alone, creating a characteristic "fried egg" appearance with a thin precursor film surrounding the bulk droplet. This spreading behavior results in a very thin liquid film with a large wetted footprint, which promotes rapid vaporization and high mean heat flux values even at moderate superheats. The thin-film geometry also means that the transition from conduction-dominated evaporation through nucleate boiling to the Partial Dryout Vapor Recoil regime occurs over a relatively compact superheat range (10–50°C), concentrating all four vaporization regimes within a single experimental platform.

A further advantage of the droplet deposition configuration is the exceptional optical access it provides. Because the droplet is deposited onto a horizontal upward-facing surface and vaporizes as a thin film, a high-speed camera (Edgertronic, 1200 fps) mounted directly above the surface captures an unobstructed top-down view of the entire wetted footprint throughout the vaporization process. This perspective reveals the full spatial distribution of nucleation sites, bubble formation and coalescence events, liquid film recession, and dryout patterns as they evolve in time. In conventional pool boiling experiments, optical access is frequently obstructed by the liquid reservoir above the heated surface, by steam rising from the boiling region, or by the presence of neighboring bubbles that occlude the camera's line of sight, particularly at higher heat fluxes. The deposited droplet configuration avoids these limitations entirely: the droplet's thin profile and open-air environment above the surface ensure that the camera captures high-fidelity images of the two-phase morphology at every stage of the vaporization event, from the initial spreading through complete evaporation. This unobstructed imaging is critical for the CNN analysis developed in this chapter, as the model's ability to extract meaningful morphological features depends directly on the quality and completeness of the visual data.

This concentration of diverse morphological states combined with high-quality optical access is what makes the dataset particularly valuable for machine learning: a single set of experiments spanning 428 data points captures the full spectrum of boiling phenomenology, from gentle film evaporation to violent vapor recoil ejection, providing the CNN with a rich diversity of visual patterns and corresponding heat transfer outcomes to learn from. By contrast, pool boiling experiments at comparable superheat ranges typically remain within one or two regimes, and their optical access is often compromised at the higher heat flux conditions that produce the most complex and information-rich morphologies.

To measure the surface temperature during these vaporization events, two K-type thermocouples are inserted into holes drilled on the side of the cylinder, approximately 1 mm from the top surface, positioned opposite each other. These thermocouples provide the surface temperature measurements used to determine the wall superheat temperature $\Delta T = T_w - T_{sat}$, where T_w is the measured wall temperature and T_{sat} is the saturation temperature of water at the system pressure (atmospheric, approximately 100°C). A high-speed camera (Edgertronic) positioned above the heated surface captures the droplet spreading and vaporization process at 1200 frames per second (fps). The camera provides a top-down view of the droplet as it spreads on the superheated surface, undergoes nucleation and vaporization, and ultimately evaporates completely.

During experiments, water droplets of controlled volume (5–8 μL, measured using a calibrated pipette at the time of deposition) were deposited on the superheated nanostructured surface at superheat temperatures ranging from 10°C to 50°C. The vaporization process was recorded using the high-speed camera, and the surface temperature was recorded simultaneously via the thermocouples connected to a data acquisition system. The experiments track the following as input data: the high-speed video images, the superheat temperature, the droplet size (initial volume), and the surface contact angle. The outputs tracked are the regime classification (conduction, isolated bubble, vigorous nucleation, or PDVR) and the mean heat flux during vaporization.

The mean heat flux during droplet vaporization was calculated using the following equation, which accounts for only the latent heat required to transform the liquid droplet into steam (the sensible heat required to raise the droplet temperature to saturation is negligible compared to the latent heat):

$$\overline{q"} = \frac{\rho V_0 h_{lv} f}{nA} \tag{3.1}$$

In Equation 3.1, $\overline{q"}$ is the mean heat flux [W/cm^2], ρ is the density of liquid water [kg/m^3], V_0 is the initial volume of the droplet [m^3], h_{lv} is the latent heat of vaporization of water [J/kg], f is the frame rate of the camera [fps], A is the average wetted area of the droplet footprint during vaporization [cm^2], and n is the number of frames between the droplet making contact with the surface and its complete evaporation. The frame rate of the camera used in this study was 1200 frames per second.

The maximum droplet wetted area was measured using high-speed videos of the droplet vaporization experiments for all test cases. Each video was stepped through until the frame where the maximum spread area was identified. The average wetted area is taken as one-half the maximum wetted area, based on the assumption that the wetted area decreases approximately linearly from its maximum value to zero as the vaporization process completes. Post-processing for the evaporation experiments was performed by importing each video, identifying the frame at which the droplet first contacts the substrate (defining time zero), and the frame at which the droplet has completely vaporized. The evaporation time was then calculated by dividing the frame count by the camera frame rate. For example, a droplet observed to vaporize completely at a video time corresponding to 600 frames would have an evaporation time of 600/1200 = 0.5 seconds (1200 fps was used for the experiments in this study). A total of 428 images were collected across the full range of superheat conditions and droplet sizes. The experimental parameter ranges are summarized in Table 3.1.

Table 3.1. Experimental parameters and uncertainty ranges for the CNN image-digital data study.

Parameter	Range	Uncertainty
Superheat Temperature (ΔT)	10–50°C	± 2°C
Droplet Size	5–8 μL	± 0.3 μL
Contact Angle	4.41–5.08°	± 0.3°
Mean Heat Flux	17.7 – 156.4 W/cm^2	± 5% (1.0 to 9.0 W/cm^2)

3.3.1. Image Data Collection

Since the coloration of the image does not carry a physical interpretation in this experiment (the droplet and surface appear in grayscale under the experimental lighting conditions), all original images were converted to grayscale. This conversion significantly reduces the computational cost for training the model since it deals with only a single-channel array representation of the image rather than three color channels. The images were then rescaled to 100×100 pixels to further reduce computational cost. At this resolution, most of the important morphological features of the image, including bubble sizes, liquid film extent, nucleation site locations, and dryout regions, are still preserved.

The data were split into two sets: 15% of the image data (65 images) was retained as a test set to evaluate the accuracy of the model on unseen data, and the remaining 85% (363 images) was used for training.

The training data were augmented using a set of augmentation parameters to improve model robustness and prevent overfitting. The augmentation parameters included random rotation of up to 5 degrees, horizontal and vertical shifts of up to 10%, shear of up to 10%, zoom of up to 10%, and random horizontal flips. Empty spaces created by the image modifications were filled using the nearest pixel value strategy. The purpose of the data augmentation is to help generate a robust model by deliberately adding controlled perturbation to the image data, as discussed by Shorten and Khoshgoftaar [57]. The data augmentation resulted in a set of 9,438 images augmented from the 363 original training images.

All input data were preprocessed to eliminate unwanted bias toward higher-magnitude values. The images were normalized by dividing all pixel values by 255, ensuring that the arrays representing images contain only values in the range between 0 and 1. The superheat temperature, droplet size, and contact angle data were normalized by dividing by their respective median values. The regime morphology classification was converted to a numerical representation via one-hot encoding, where each regime class is represented by a binary vector of 4 bits with one of the bits containing the value 1, indicating the class, and all other elements set to 0.

3.4. CNN Architectures for Experimental Data

Three CNN architectures were developed to progressively explore the value of incorporating different data modalities. The three cases are summarized in Figure 3.10. Case A examines the architecture's strength with information unrelated to the appearance of the droplet (digital data only). Case B examines the architecture with information unrelated to the temperature of the system (image + non-thermal digital data). Case C examines the architecture considering both the appearance of the droplet and the temperature of the system (image + all digital data). As this development involved a large amount of data (more than 9000 images with a 100x100 size), a computationally efficient activation function such as ReLU was needed in the solution implementation. Unlike sigmoid or tanh functions, ReLU is simple to compute with no exponentials or division. Also, unlike sigmoid or tanh, which saturate and produce tiny gradients for extreme inputs, ReLU avoids vanishing gradients for positive inputs, helping deep networks train more effectively [58].

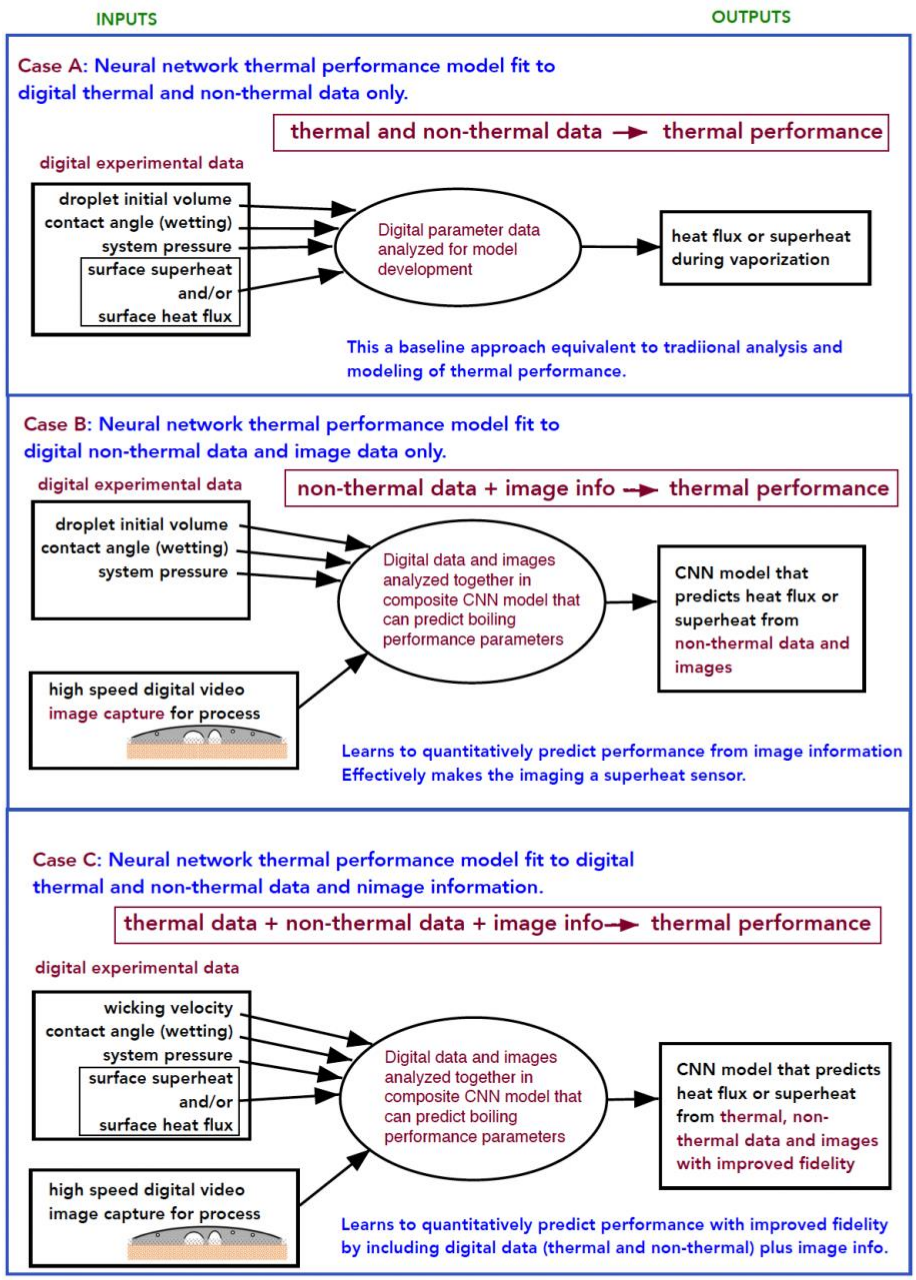


Figure 3.10. Model cases considered in this study, highlighting the inputs and outputs

3.4.1. Case A: Digital-Only Baseline

Case A was designed to establish a baseline for prediction accuracy using only digital data without any image information. The inputs consist of the surface superheat temperature ΔT, the initial droplet volume V_0, and the contact angle θ. The architecture is a simple fully connected neural network. In the initial configuration with a single dense layer of 32 neurons, the mean heat flux prediction achieved a RMSPE of 13.2% on the training data and 13.6% on the testing data [5].

To ensure a fair comparison with the more complex architectures in Cases B and C, Case A was re-evaluated with 64 neurons in the dense layer (matching the architecture used in the subsequent cases).

This configuration yielded a RMSPE of 12.9% on training data and 15.8% on testing data. The slight degradation in test performance with more neurons may indicate mild overfitting, but more importantly, the results confirm that digital-only models plateau at approximately 13–16% RMSPE regardless of network complexity. This plateau establishes the fundamental limitation of digital-data-only approaches and motivates the incorporation of image data in Cases B and C. The Case A architecture and resulting prediction are shown in

Figure 3.11 and Figure 3.12. Also, Figure 3.13 shows a surface plot generated to predict the mean heat flux for Case A. The contact angle for the plot was fixed at 4.9°. This plot spans the range of values for the droplet sizes and temperature values studied in this paper. It provides a full picture of the predictive model behavior throughout the droplet sizes and superheat temperature space. From the surface plot, we can observe that the droplet size has limited effects on the mean heat flux at low superheat, but the effect gradually increases at higher superheat values.

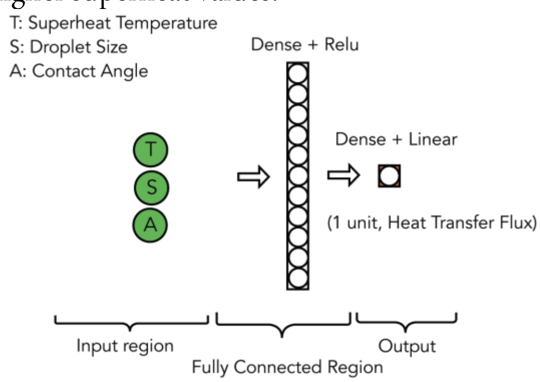


Figure 3.11. Artificial Neural Network for Case A.

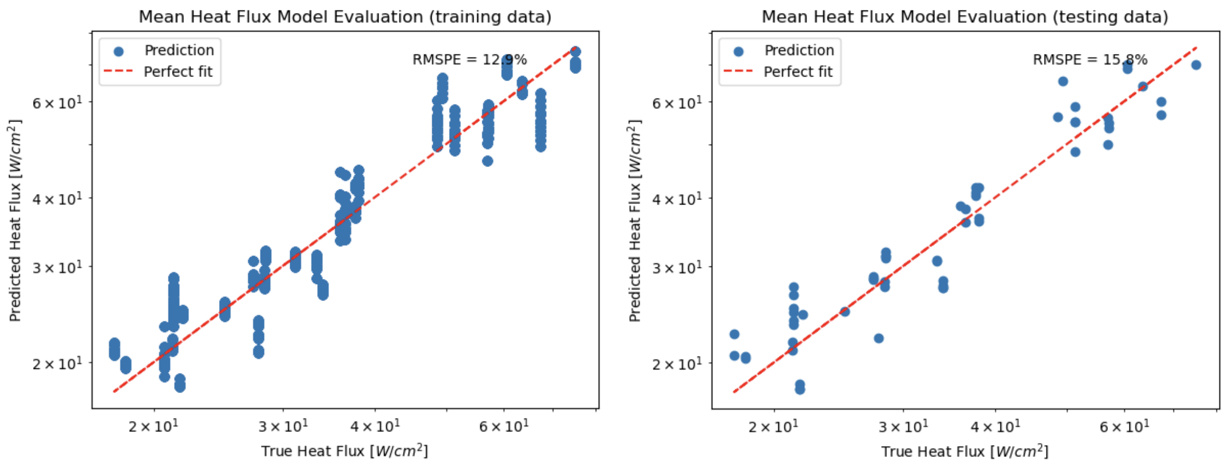


Figure 3.12. Mean heat flux rate prediction versus data results for Case A on (a) the training dataset and (b) the testing (validation) dataset.

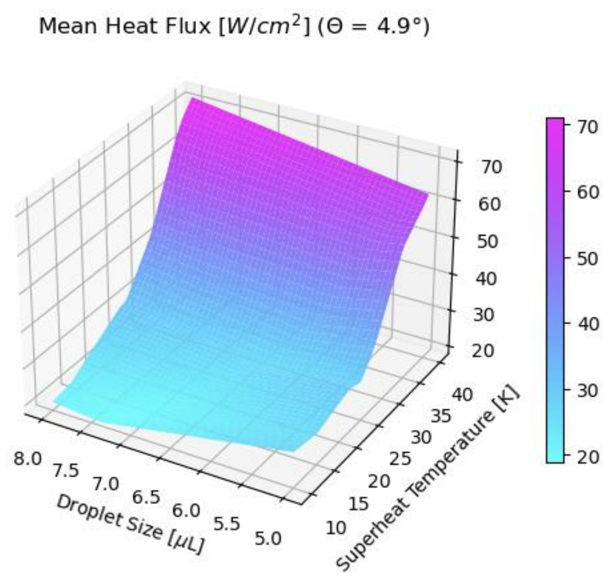


Figure 3.13. Surface plot prediction of mean heat flux during droplet vaporization for Case A

3.4.2. Case B: Image with Non-Thermal Digital Data

Case B was designed to evaluate how well a model can predict thermal performance using only droplet size, contact angle, and image information, explicitly excluding direct temperature measurement. This case tests the hypothesis that the CNN can learn to associate visual morphological features with the thermal state of the system, essentially functioning as a non-contact temperature sensor. The architecture, depicted in Figure 3.14, takes a 100×100 grayscale array representation of the image, the droplet size variable, and the contact angle variable as input, generating both the wall superheat temperature and mean heat flux as outputs.

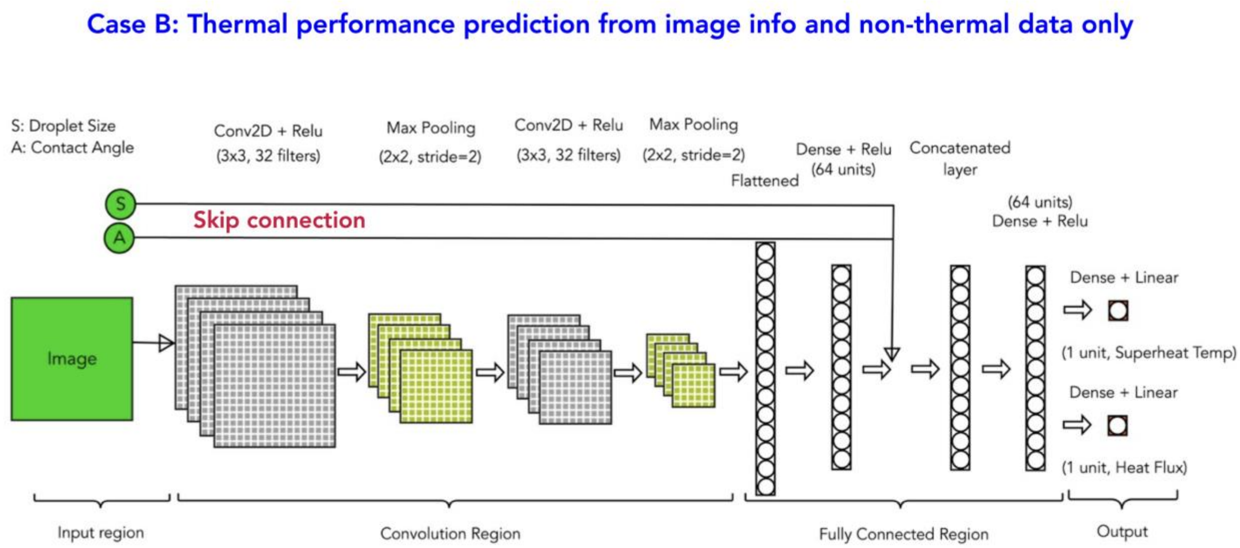


Figure 3.14. Convolution neural network (CNN) architecture for Case B

The image data passes through two sequential convolution blocks. The first block consists of a 3×3 convolution layer with 32 filters and same padding, followed by ReLU (Rectified Linear Unit) activation to introduce non-linearity, and 2×2 max pooling with a stride of 2 and same padding to summarize the features captured by the convolution layer. This set of three operations is repeated one more time in the second convolution block. The convolution operations extract increasingly abstract spatial features from the image: the first layer captures low-level features such as edges, gradients, and local texture patterns, while the second layer captures higher-level combinations of these features that correspond to more complex morphological structures such as bubble boundaries, wetting fronts, and dryout regions.

As the data enters the fully connected region, the output from the convolution region is flattened into a one-dimensional vector. This flattened vector passes through a dense layer with 64 neurons and ReLU activation. A skip connection concatenates the droplet size and contact angle inputs directly to the dense layer output, bypassing the convolutional processing. This skip connection serves two purposes: it provides the non-thermal digital data directly to the fully connected layers without requiring them to be recovered from the image features, and it allows the network to learn how these variables interact with the image-derived features. The concatenated data flows through another dense layer with 64 neurons and ReLU activation before connecting to two final output dense layers, each with one neuron: one outputting the superheat temperature and one outputting the mean heat flux. The predictions are shown in Figure 3.15 and Figure 3.16

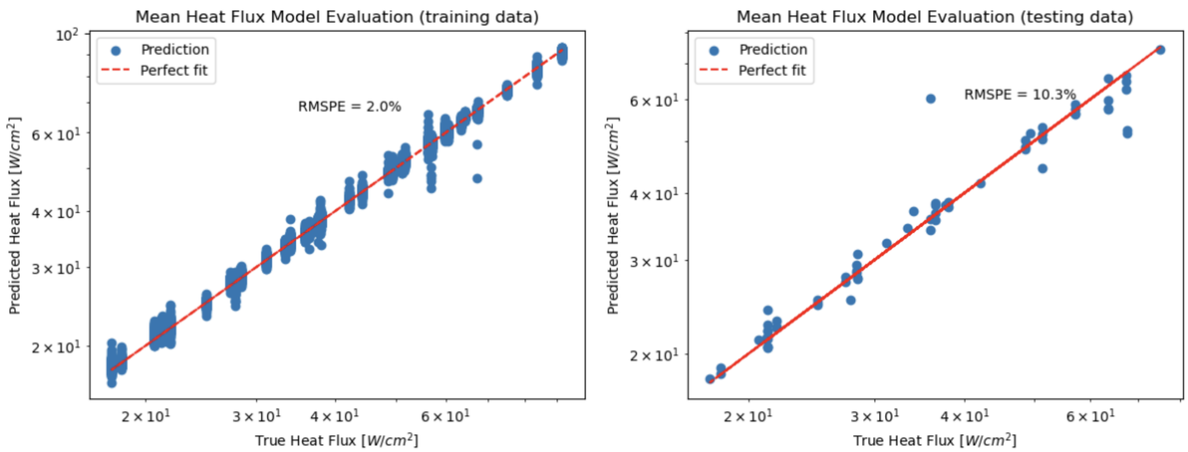


Figure 3.15. Mean heat flux prediction results for Case B on (a) the training dataset and (b) the testing dataset.

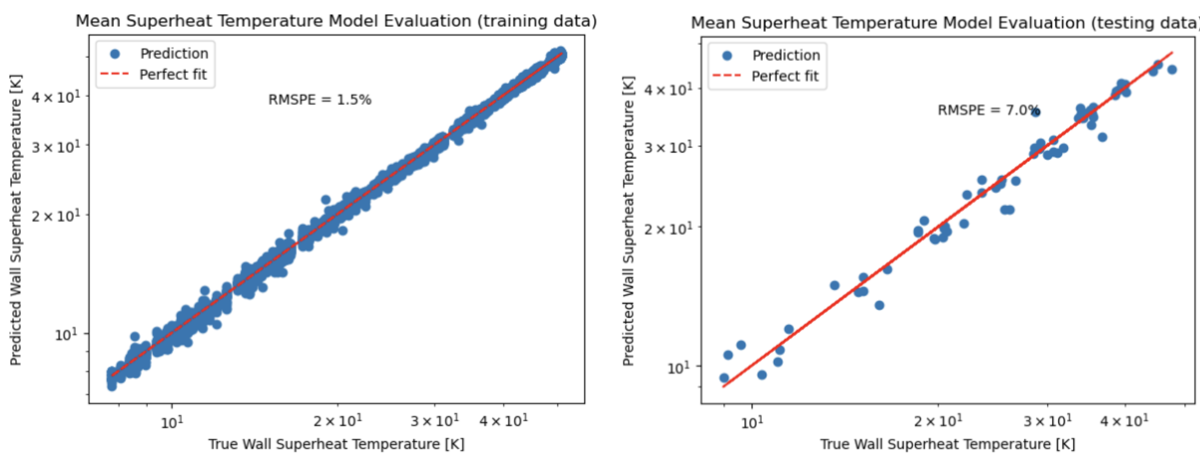


Figure 3.16. Superheat temperature difference prediction for Case B on (a) the training and (b) the testing datasets.

Case B achieved a mean heat flux prediction RMSPE of 2.0% on training data and 10.3% on testing data, with superheat temperature prediction of 1.5% on training and 7.0% on testing data. We explored fitting our Case B CNN model with both 32 neurons and 64 neurons in the fully connected regions. Increasing to 64 neurons in the fully connected region improved the fit to the experimental data slightly (by only about 3–4%), indicating diminishing returns. Since the results for 64 neurons provided very good fits and increasing the number of neurons further would increase the probability of overfitting, we accepted the 64-neuron results as the optimal compromise for the CNN design.

The ability to predict superheat temperature from image information alone (7.0% RMSPE) demonstrates that the CNN learns to associate visual morphological features with thermal state, the shape, spread, and bubble activity visible in the image encode sufficient information to infer the surface temperature within approximately 7% error. This is a significant finding because it demonstrates that visual morphology and thermal state are strongly (though not perfectly) correlated, and that a trained CNN model can function as a non-contact temperature sensor.

3.4.3. Case C: Image with Full Digital Data

Case C combined image data with all available digital inputs including superheat temperature to predict the nucleation regime (four classes: conduction, isolated bubble, vigorous nucleation, PDVR) and mean heat flux. The architecture, shown in Figure 3.17, employs the same convolution and dense layer structure as Case B, with the critical addition of the superheat temperature as a skip connection input alongside droplet size and contact angle.

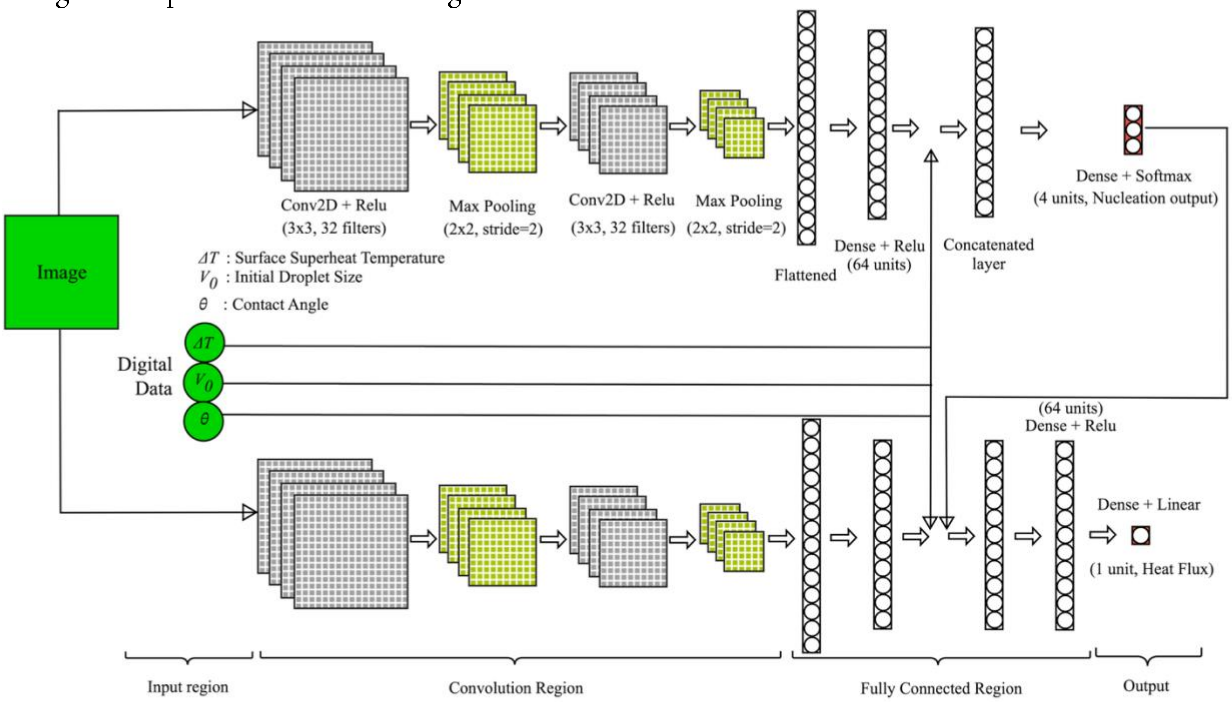


Figure 3.17. CNN architecture for Case C.

The CNN architecture consists of four inputs: a 100×100 grayscale array representing the image, wall superheat temperature, droplet size, and contact angle. The output structure consists of a nucleation regime classification with four classes and a continuous mean heat flux prediction. Unlike Case B, which uses a single convolutional path, Case C employs a hybrid parallel-series design with two independent convolutional branches processing the same image. Each branch consists of two convolution blocks (3×3 filters with 32 filters per layer, ReLU activation, and 2×2 max pooling), followed by flattening and a Dense layer with 64 neurons. The first convolutional branch feeds the regime classification output: its Dense layer output is concatenated with the superheat temperature, droplet size, and contact angle via a skip connection, and the concatenated vector passes through a softmax layer to produce the four-class regime prediction. The second convolutional branch feeds the heat flux prediction in a cascaded design: the regime classification output (a four-element probability vector) is concatenated with the second branch's Dense layer output and the same skip-connected digital inputs, then passed through a Dense layer with 64 neurons and ReLU activation before the final linear output for heat flux. This cascaded architecture allows the two branches to learn different visual features optimized for their respective tasks, while the series connection enables the predicted regime to inform the heat flux prediction, reflecting the physical reality that heat transfer mechanisms differ qualitatively across boiling regimes. Case C achieved a mean heat flux prediction RMSPE of 1.0% on training data and 5.8% on testing data, as shown in

Figure 3.18. This represents a 57.4% relative reduction in heat flux prediction error compared to the digital-only baseline (from 13.6% to 5.8% RMSPE for Case A vs. Case C using the 32-neuron baseline, or from 15.8% to 5.8% using the 64-neuron version). For regime classification, the model achieved 100% accuracy on the training data and misclassified only 2 out of 65 images on the testing data (96.9% accuracy). The two misclassified images occurred at the boundary between the vigorous nucleation and PDVR regimes, where the visual morphology is most ambiguous and transitional features of both regimes may coexist.

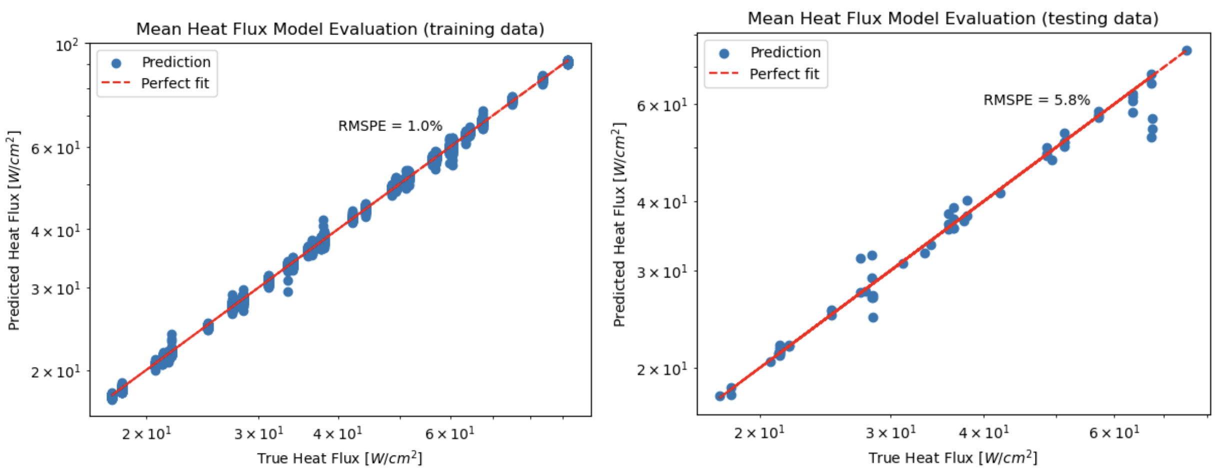


Figure 3.18. Mean heat flux prediction results for Case C on (a) the training and (b) the testing datasets.

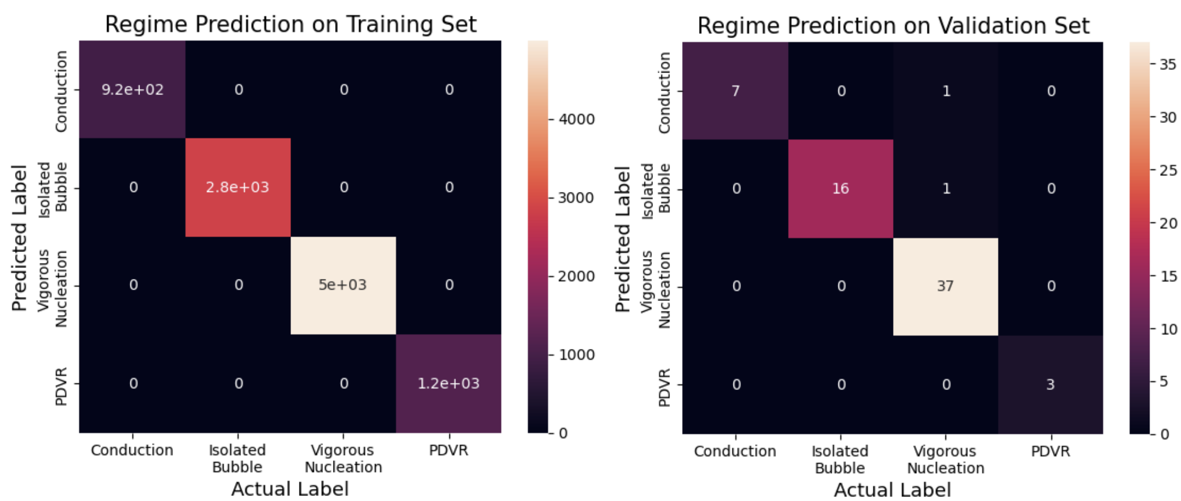


Figure 3.19. Nucleation regime prediction for Case C for (a) the training dataset and (b) the testing dataset.

3.5. Results and Discussion

The progressive improvement across the three cases provides compelling evidence that image data contains substantial quantitative information about the boiling process that complements traditional digital measurements. The key results are summarized in Table 3.2.

Table 3.2. Comparison of prediction performance across CNN architectures.

Architecture	Input Modalities	Heat Flux RMSPE (Train)	Heat Flux RMSPE (Test)
Case A	Digital Only	12.9%	15.8%
Case B	Image + non-thermal	2.0%	10.3%
Case C	Image + all digital	1%	5.8%

The 5.8% RMSPE achieved by Case C is comparable to the best fits for quality nucleate boiling data reported in the literature, suggesting that the CNN solution approaches the fundamental accuracy limits imposed by experimental uncertainty and the inherent stochasticity of the boiling process. The predictions are physically realistic, with continuous heat flux variation across regime boundaries rather than artificial discontinuities, as shown in

Figure 3.20.

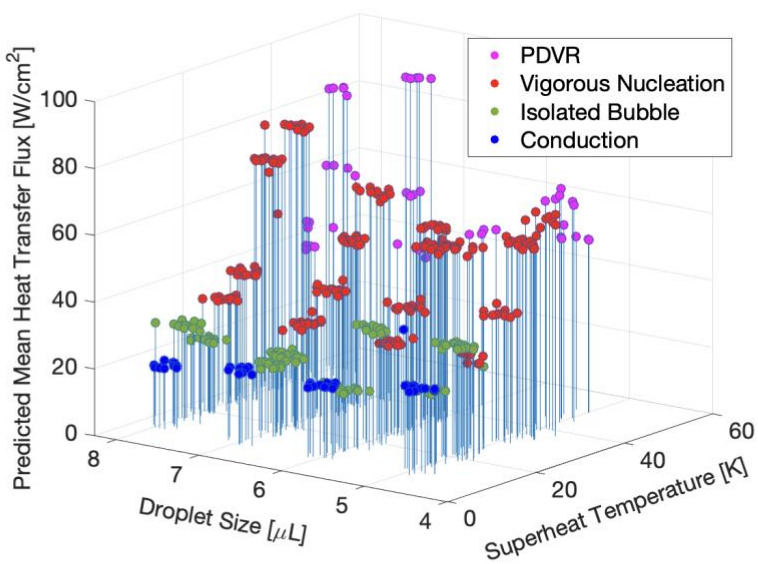


Figure 3.20. Predicted mean heat flux using Case C model over the entire dataset.

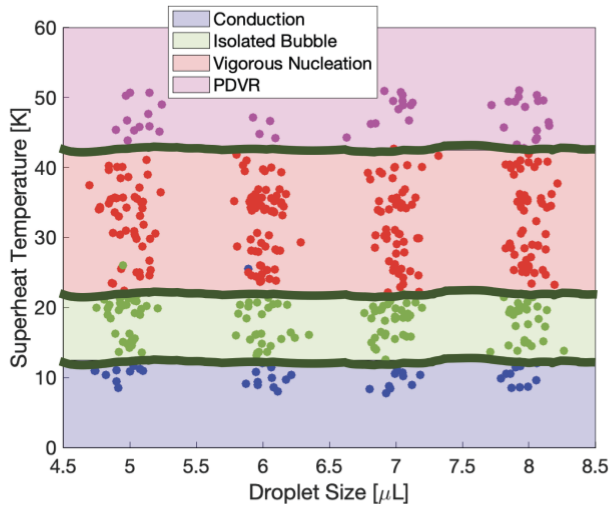


Figure 3.21. A top view of Figure 3.20 clearly showing the data points distribution over the different regimes.

The stem plot in Figure 3.20 reveals that the difference in mean heat flux along the droplet size axis becomes more significant as the superheat temperature increases, which is physically consistent with the increasing importance of nucleation and vapor recoil at higher superheats. This pattern was also observed in the surface plot for Case A, confirming that the CNN model learns physically meaningful relationships between the input variables and the heat transfer performance.

The architecture with 64 neurons improved fits by 3–4% over 32 neurons, though with diminishing returns, suggesting the model complexity is well-matched to the available data.

Significantly, the CNN model using the Python Keras functional API successfully combined convolution layers for image processing with fully connected dense layers for digital data processing, producing a high-performing model capable of simultaneously predicting the mean heat flux, superheat temperature, and morphology regime as needed.

3.6. CNN as a Temperature Sensor

An important secondary finding of this work is that a trained CNN can function as a temperature sensor. To further explore this capability, a dedicated CNN model was trained to predict superheat temperature from image data alone, without any digital data inputs. The architecture, shown in Figure 3.22, is similar to Case B but configured with only the 100×100 grayscale image as input and the wall superheat temperature as the sole output.

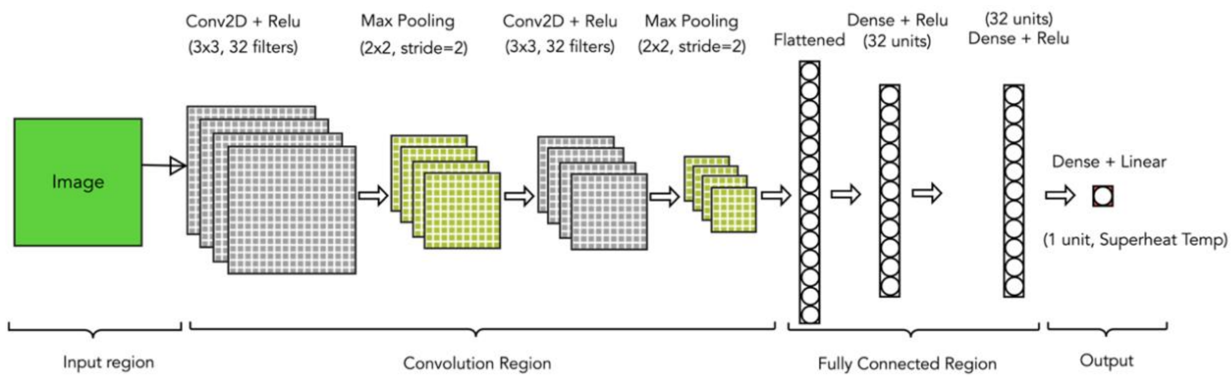


Figure 3.22. CNN architecture for predicting the superheat temperature difference from image information only.

The image data passes through two convolution blocks identical to those in Cases B and C (3×3 convolution with 32 filters, ReLU activation, 2×2 max pooling in each block). The output from the convolution region is flattened and passed through two dense layers with 32 neurons each (not 64, as this simpler architecture was found to be sufficient for single-output prediction). The final output is a single neuron predicting the superheat temperature.

The image-only CNN achieved a superheat prediction RMSPE of 10.9% on the testing data. When this predicted superheat is subsequently used as input to the Case C model structure (cascading the image-predicted temperature into the digital data input), the resulting heat flux prediction achieves a RMSPE of 14%. While this cascaded approach does not match the performance of Case C with directly measured temperature data, it demonstrates the practical potential for CNN-based temperature sensing in environments where direct temperature measurement is difficult or impossible. This capability could be valuable in industrial boiling systems where thermocouple placement near the boiling surface is impractical due to geometric constraints, corrosive environments, or the need for non-intrusive monitoring.

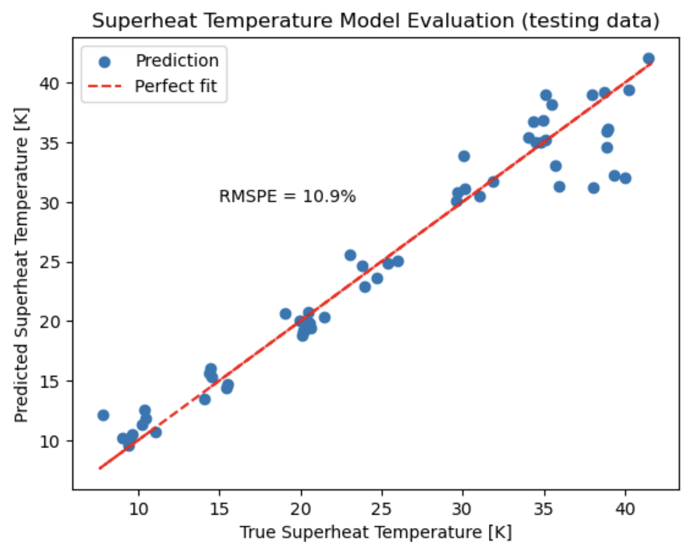


Figure 3.23. Superheat temperature difference prediction for the CNN model as a temperature sensor.

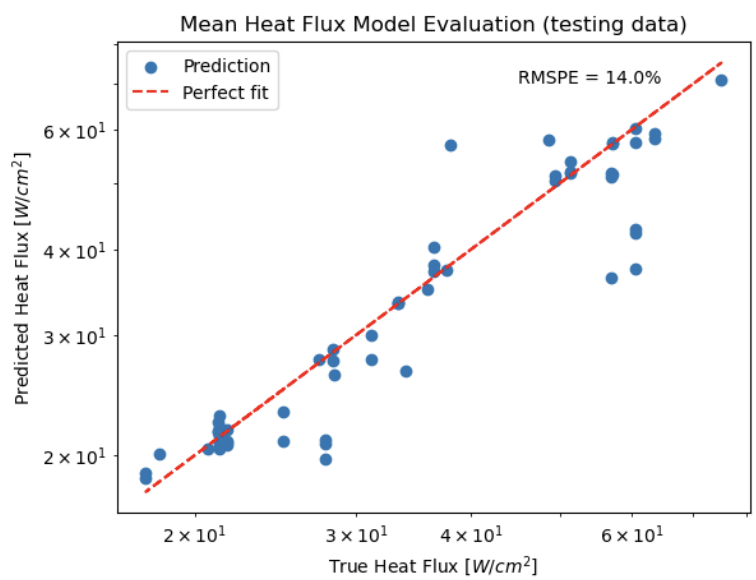


Figure 3.24. Mean heat flux prediction using the temperature predicted from a CNN (model B) as input to model A.

3.7. Conclusion

The work presented in this chapter demonstrates that convolutional neural networks can effectively integrate heterogeneous data modalities, high-speed video imagery and digital sensor measurements, to achieve substantially improved prediction of boiling heat transfer performance on ZnO nanostructured surfaces. The progressive improvement from 15.8% (Case A, digital only) to 10.3% (Case B, image + non-thermal) to 5.8% (Case C, image + all digital) RMSPE confirms that each data modality contributes unique, complementary information about the boiling process. The hybrid parallel-series CNN architecture with skip connections proves effective for simultaneously processing high-dimensional image data through convolutional layers and low-dimensional digital data through dense layers, with the skip connection enabling direct interaction between these two information streams.

Historically, the analysis of boiling heat transfer has relied on physics-based correlations fitted to digital sensor data, with visual observation used primarily for qualitative regime identification rather than quantitative prediction. The results of this chapter represent a departure from that paradigm: the CNN treats visual morphology as a quantitative input on equal footing with traditional measurements, and the improvement in prediction accuracy demonstrates that substantial information content in the images was previously untapped by conventional analysis methods.

The results also demonstrate that the visual morphology of the two-phase mixture encodes sufficient thermal state information for a CNN to function as a non-contact temperature sensor, achieving 10.9% RMSPE in superheat prediction from images alone. This capability has practical implications for industrial boiling systems where direct thermocouple placement near the boiling surface is impractical.

Despite these advances, the 5.8% RMSPE achieved by Case C suggests that some information about the boiling process remains uncaptured by the combination of imaging and digital sensors. The boiling process produces acoustic emissions, the sounds of bubble nucleation, growth, coalescence, and collapse, that carry temporal dynamic information at frequencies beyond what the camera can resolve. The question of whether acoustic data can provide further complementary information to improve prediction accuracy beyond what imaging and digital sensors achieve is addressed in Chapter 4.

Chapter 4. Multimodal Sensor Fusion with Digital Data, Images, and Acoustic Emission Data

The preceding chapter demonstrated that combining image data with digital sensor measurements through a hybrid CNN architecture achieves substantially improved heat flux prediction compared to digital-only approaches. However, the boiling process generates information beyond what cameras and thermocouples can capture. Acoustic emissions produced by bubble nucleation, growth, coalescence, and collapse carry temporal dynamic information at frequencies and timescales that complement both visual and thermal measurements. This chapter extends the multimodal framework to incorporate acoustic emission data as a third sensing modality, developing the quartile-based feature extraction methodology, evaluating acoustic-only prediction capability, and demonstrating that the fusion of all three modalities achieves the best prediction performance reported in this dissertation.

4.1. Motivation: Prior research exploring acoustic characteristics of boiling processes

Acoustic emissions during boiling processes have long been recognized as carrying valuable information about the underlying thermophysical phenomena. The characteristic sounds of boiling, ranging from the gentle crackling of incipient nucleation to the violent popping of film boiling, are direct manifestations of bubble dynamics, including nucleation, growth, coalescence, and collapse. Previous studies have demonstrated that acoustic signatures can provide insights into boiling regimes, bubble departure frequencies, and heat transfer intensities [59, 60].

The physical basis for acoustic emission during boiling is well understood. When a bubble nucleates on a heated surface, its rapid growth creates a pressure pulse in the surrounding liquid. When bubbles depart from the surface or collapse during condensation in subcooled regions, additional pressure pulses are generated. The frequency content of these acoustic emissions is governed by the natural frequency of oscillation of the bubbles, which depends on bubble size, liquid properties, and ambient pressure. For the bubble sizes typical of nucleate boiling at atmospheric pressure (diameter ~0.1–5 mm), the characteristic frequencies fall in the range of hundreds of Hz to several kHz, well within the range of conventional acoustic sensors.

Sinha et al. [61] demonstrated the power of acoustic-based machine learning for boiling analysis by training a CNN on acoustic spectrograms for boiling regime classification, achieving 99.92% validation accuracy. Their study used a hydrophone sampling at 10 kHz, with data divided into one-second audio files converted to spectrograms. The dataset comprised of 2,056 background samples, 13,367 nucleate boiling samples, 399 pre-CHF samples, and 460 transition boiling samples. Their analysis revealed that peak acoustic frequency shifts systematically from 200–250 Hz during nucleate boiling to 400–500 Hz at CHF onset, providing a clear acoustic signature of the approaching boiling crisis.

Dunlap et al. [62] investigated nonintrusive heat flux quantification using acoustic emissions during pool boiling, testing six machine learning architectures and achieving their best performance with a Fast Fourier Transform combined with Gaussian Process Regression (FFT-GPR), yielding a mean absolute percentage error (MAPE) of 4.5% and R^2 of 0.999. Their analysis demonstrated that acoustic features below 512 Hz are most significant for heat flux prediction, and that prediction performance improves with increasing temporal length of acoustic sequences.

Lim et al. [63] compared four deep learning architectures, DNN, CNN, Transformer, and Fourier Neural Operator (FNO), for predicting boiling heat flux, heat transfer coefficient, and regime from externally mounted acoustic emission sensors. Their study found that the Transformer architecture with spectrogram input achieved the best overall performance, with less than ±20% error margins for heat flux and HTC prediction and 97.6% accuracy for boiling regime classification, while the CNN with spectrogram input achieved the highest classification accuracy at 98.3%.

These studies collectively demonstrate that acoustic emissions contain substantial information about boiling processes, but they have focused on single-modality (acoustic-only) analysis. The investigation presented in this chapter, based on our submission to the ASME 2026 Heat Transfer Summer Conference [19], extends our previous multimodal framework to incorporate acoustic data as a third sensing modality alongside imaging and digital sensors, assessing whether the combination of all three modalities can achieve prediction accuracy beyond what any single or dual modality can provide.

4.2. Experimental Methods and Acoustic Data Collection

The experimental apparatus used in this study builds upon the droplet deposition experiment described in Chapter 3, with the addition of acoustic sensing instrumentation. The experimental setup is depicted in Figure 4.1. The key modification is the addition of a RØDE Wireless ME microphone system, a compact wireless microphone system employing a pre-polarized pressure transducer (condenser) with an omnidirectional polar pattern. The transmitter unit, positioned approximately 3 cm from the heated surface, features a built-in condenser microphone capsule with a frequency response of 20 Hz to 20 kHz, a maximum sound pressure level of 122 dB SPL, an equivalent noise level of 22 dBA (A-weighted), and a dynamic range of 100 dB. The omnidirectional pickup pattern ensures that acoustic emissions from all regions of the droplet footprint are captured without directional bias, which is important given that nucleation events can occur at any location on the heated surface. The transmitter communicates wirelessly with the receiver unit via RØDE's Series IV 2.4 GHz digital transmission with 128-bit encryption, providing stable audio transmission at distances up to 200 m with near-zero latency. The receiver was connected to a computer via USB-C, recording the acoustic signal at 24-bit resolution and a sampling rate of 48 kHz. This sampling rate provides a Nyquist frequency of 24 kHz, which offers substantial headroom above the frequencies of interest for bubble dynamics, which typically fall in the range of hundreds of Hz to several kHz at atmospheric pressure.

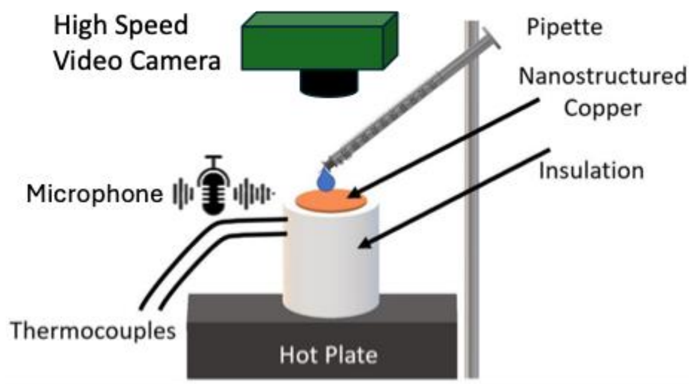


Figure 4.1. Schematic of the droplet deposition/vaporization experiment with acoustic sensing system.

The experiment documents five types of input data: images from the high-speed camera, acoustic waveforms from the microphone, superheat temperature from thermocouples, droplet size from pipette calibration, and contact angle from surface characterization. The outputs are the regime classification (four classes) and the mean heat flux calculated using Equation (3.1).

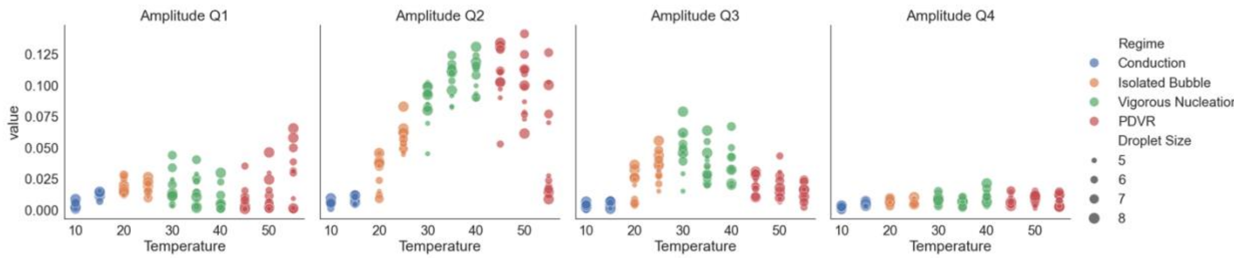
The expanded dataset for this study comprises 452 high-speed video images captured from 123 unique droplet vaporization events across the full superheat range (10–55°C), with each vaporization event producing multiple video frames but a single acoustic recording from which one set of 12 quartile features is extracted. The data split follows the same protocol as Chapter 3: 15% of the data (80 images and corresponding acoustic recordings) was retained as the test set, with the remaining 85% used for training. The image data were processed identically to the procedure described in Section 3.3.1 (grayscale conversion, 100×100 pixel rescaling, normalization, augmentation). The experimental parameters and their ranges are summarized in Table 4.1.

Table 4.1. Range of experimental parameters and data collected.

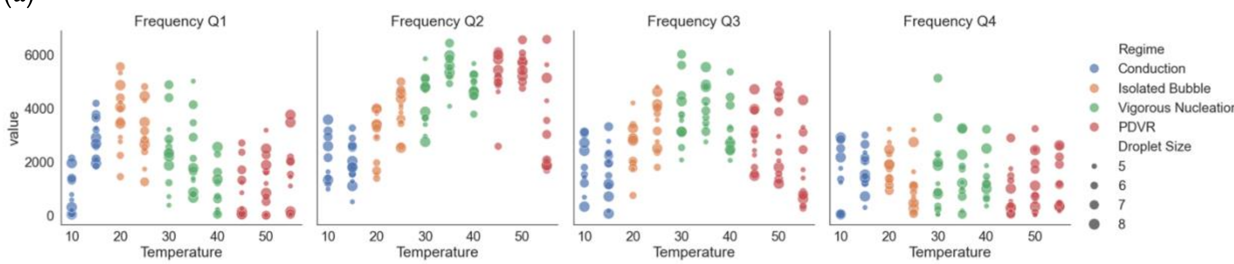
Parameter	Range	Uncertainty
Superheat Temperature (ΔT)	10 – 55°C	± 2 °C
Droplet Size	5 – 8 μL	± 0.3 μL
Contact Angle	6°	± 0.3°
Mean Heat Flux	12.1 – 134.1 W/cm²	± 5 [%], 1.0 to 9.0 W/cm²
Acoustic Sampling Rate	48 kHz	
Video Frame Rate	1200 fps	
Total Image Data Points	452 images (80 testing)	

4.3. Acoustic Signal Characteristics and Feature Extraction

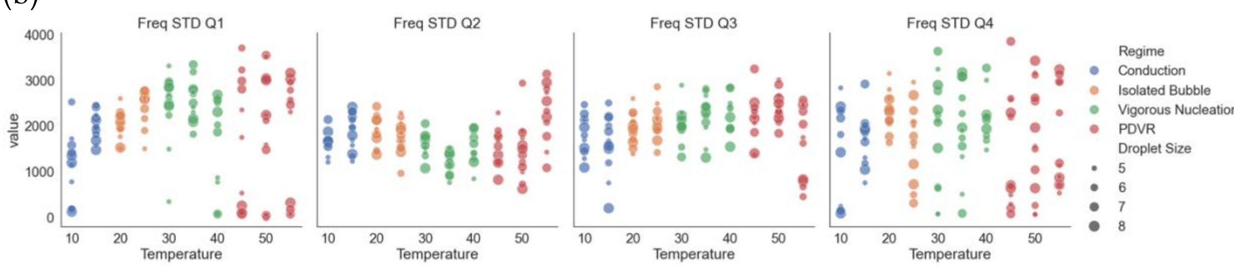
The acoustic emissions during droplet vaporization exhibit distinct characteristics that vary systematically with the boiling regime, as illustrated in Figure 4.2, which presents representative acoustic waveforms alongside corresponding images for each of the four regimes. The RØDE Wireless ME microphone captures acoustic emissions across its full frequency response of 20 Hz to 20 kHz, with the 48 kHz sampling rate providing a Nyquist frequency of 24 kHz that offers substantial headroom above the frequencies of interest for bubble dynamics. Analysis of the quartile-based acoustic features extracted from 123 experimental test cases reveals quantitatively distinct acoustic signatures for each regime, summarized in Table 4.2.



(a)



(b)



(c)

Figure 4.2. Quartile-based acoustic features across the superheat range for all four boiling regimes: (a) mean amplitude, (b) mean peak frequency, and (c) peak frequency standard deviation. Marker color indicates regime classification; marker size indicates initial droplet volume in μL

Table 4.2. Acoustic feature characterization by boiling regime, computed from 123 experimental test cases. Q2 denotes the second temporal quartile of the vaporization event, which corresponds to peak nucleation activity across all regimes.

Regime	ΔT Range (°C)	n	Mean Amp	Max Amp	Q2 Mean Amp	Mean Freq (Hz)	Q2 Mean Freq (Hz)	Mean Freq STD (Hz)
Conduction	10–15	24	0.0058	0.0156	0.0069	1,806	1,955	1,657
Isolated Bubble	20–25	26	0.0242	0.0829	0.0443	2,821	3,414	2,009
Vigorous Nucleation	30–40	36	0.0402	0.1308	0.0991	2,982	4,818	1,970
PDVR	45–55	37	0.0318	0.1412	0.0862	2,352	4,667	1,815

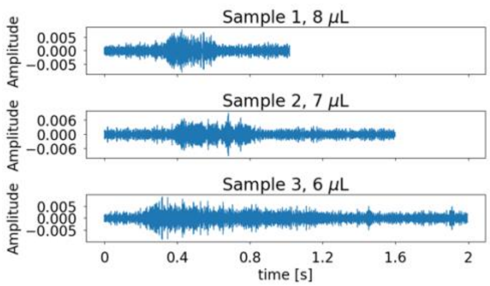
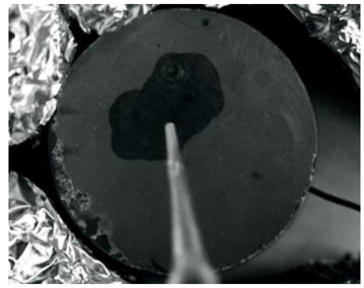
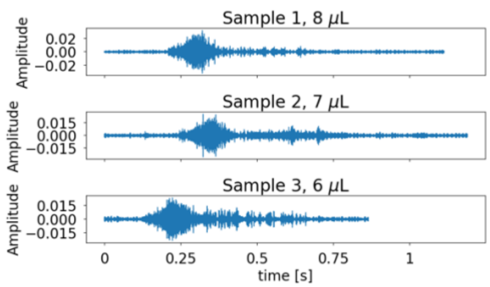
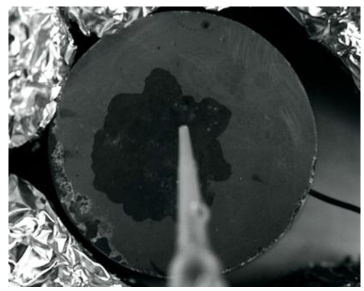
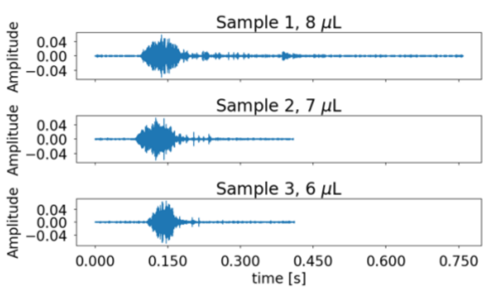
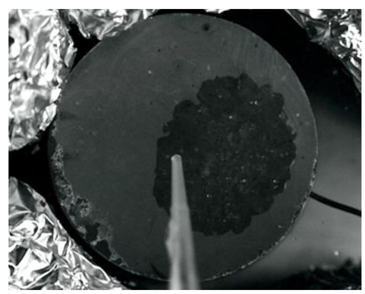
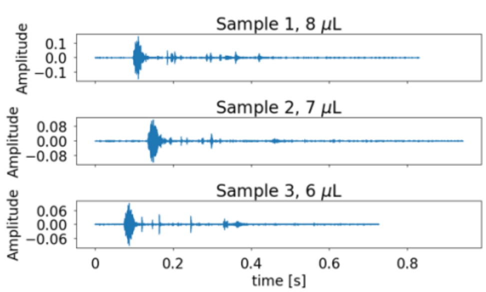
Regime	Audio waveform	image	Superheat Temperature (°C)	Evaporation time (s)
(a) Conduction		 8 μL	10	S1 = 1.56 S2 = 1.67 S3 = 1.51
(b) Isolated Bubble		 8 μL	20	S1 = 0.55 S2 = 0.62 S3 = 0.7
(c) Vigorous Nucleation		 8 μL	30	S1 = 0.33 S2 = 0.3 S3 = 0.22
(d) PDVR		 8 μL	50	S1 = 0.64 S2 = 0.51 S3 = 0.88

Figure 4.3. Representative data collected from experiments, including acoustic waveforms showing characteristic patterns for (a) conduction-dominated, (b) isolated bubble, (c) vigorous nucleation, and (d) PDVR regimes.

In the conduction-dominated regime (ΔT = 10–15°C, n = 24 test cases), the acoustic signal is characterized by low-amplitude emissions with a mean amplitude of 0.0058 and a maximum of 0.0156 across all quartiles, as shown in
Figure 4.2 (a). The mean frequency content is centered around 1,806 Hz (range: 46–3,577 Hz), with a Q2 mean frequency of 1,955 Hz, as shown in
Figure 4.2 (b), reflecting the gentle, broadband noise generated by quiescent evaporation at the liquid–air interface with minimal impulsive events. Typical evaporation times in this regime are on the order of 1.5 seconds for an 8 μL droplet. Uniquely among the four regimes, the temporal evolution of amplitude across the four quartiles is monotonically decreasing (Q1 = 0.0084 → Q2 = 0.0069 → Q3 = 0.0044 → Q4 = 0.0033), with a Q2/Q1 median ratio of 0.9×. This declining profile reflects the steady nature of conduction-dominated evaporation: the liquid film is thickest and most acoustically active immediately after deposition, and the gentle evaporation gradually subsides as the liquid volume decreases without the onset of discrete nucleation events. This monotonically decreasing temporal signature is qualitatively distinct from all three nucleation regimes, providing the neural network with an unambiguous acoustic fingerprint for the conduction regime.
As isolated bubbles begin to nucleate and depart at moderate superheats (ΔT = 20–25°C, n = 26 test cases), the acoustic emissions increase substantially, with a mean amplitude of 0.0242 and a maximum of 0.0829 — approximately 4× and 5× higher than the conduction regime, respectively. The mean frequency increases to 2,821 Hz (range: 75–5,554 Hz), with the Q2 mean frequency reaching 3,414 Hz, reflecting the discrete pressure pulses generated by individual bubble nucleation, growth, and departure events. Each bubble event produces a distinct acoustic impulse, and the spacing between impulses reflects the nucleation frequency. Unlike the conduction regime, the temporal amplitude pattern now shows a clear Q2 peak (Q1 = 0.0184 → Q2 = 0.0443 → Q3 = 0.0274 → Q4 = 0.0069), with a Q2/Q1 ratio of approximately 2.7×, as nucleation activity intensifies during the second quarter of the vaporization event then diminishes as the liquid volume decreases. At 20°C superheat, the evaporation time decreases to approximately 0.55–0.70 seconds for an 8 μL droplet, reflecting the enhanced heat transfer from nucleate boiling.
The vigorous nucleation regime (ΔT = 30–40°C, n = 36 test cases) produces the most acoustically energetic signals on average, with a mean amplitude of 0.0402 and a maximum of 0.1308 — approximately 7× and 8× higher than conduction. The Q2 mean amplitude reaches 0.0991, approximately 14× the conduction-regime Q2 value, making this the single most discriminating feature between regimes, as clearly visible in
Figure 4.2 (a). The mean frequency increases further to 2,982 Hz, with the Q2 mean frequency reaching 4,818 Hz — more than double the conduction-regime Q2 frequency — reflecting the superposition of many simultaneous bubble events generating broadband, high-frequency acoustic emissions, as shown in
Figure 4.2 (b). The temporal pattern in this regime shows a dramatic ramp-up, with a Q2/Q1 ratio of approximately 21× (mean), as the nucleation process intensifies rapidly during the second quartile before decaying through Q3 and Q4 as the liquid volume diminishes. The amplitude increases from Q1 to Q2 with a steeper slope than its subsequent decrease through Q3 and Q4. The frequency standard deviation in this regime (mean: 1,970 Hz) is among the highest, as shown in

Figure 4.2 (c), reflecting the diversity of bubble sizes and the multiple types of acoustic events — nucleation, coalescence, departure — occurring simultaneously. At 30°C superheat, evaporation times are approximately 0.22–0.33 seconds for an 8 μL droplet.

The PDVR regime (ΔT = 45–55°C, n = 37 test cases) generates the most intense individual acoustic events, with the highest maximum amplitude of 0.1412 across all regimes, corresponding to violent liquid ejection and rapid vaporization caused by vapor recoil forces. However, the mean amplitude (0.0318) is lower than the vigorous nucleation regime (0.0402), reflecting the intermittent nature of acoustic emissions in this regime — high-amplitude bursts from vapor recoil ejection events are separated by periods of relative quiet when less of the surface is in liquid contact due to dryout. This contrast between high peak and moderate mean amplitude is clearly visible in

Figure 4.2 (a), where the PDVR data points span the widest vertical range of any regime. The intermittency is also reflected in the high variability of Q2 amplitude (standard deviation 0.039, compared to 0.017 for vigorous nucleation). The Q2/Q1 ratio reaches its highest values in this regime (mean: 37×), indicating the most extreme temporal contrast between the quiet initial spreading phase and the subsequent violent vaporization. The mean frequency is 2,352 Hz (range: 0–6,571 Hz), with the Q2 mean frequency at 4,667 Hz — similar to vigorous nucleation but with a broader overall distribution, as evident in

Figure 4.2 (b) and (c), reflecting the chaotic, stochastic nature of vapor recoil events. At 50°C superheat, evaporation times are approximately 0.51–0.88 seconds for an 8 μL droplet, longer than in the vigorous nucleation regime because less surface area remains wetted due to dryout and vapor recoil.

4.3.1. Quartile-Based Feature Extraction

To prepare the acoustic data for neural network analysis, we developed a quartile-based feature extraction approach. This approach was motivated by two considerations: (1) raw acoustic waveforms sampled at 48 kHz contain far more data than needed for the neural network, creating unnecessary computational burden and potential overfitting risk, and (2) the key information in the acoustic signal — amplitude level, dominant frequency content, and temporal variability — can be captured by statistical summary features computed from a time-frequency representation without retaining the full waveform.

The feature extraction proceeds as follows. First, the raw acoustic signal for each experimental test case is denoised using the noisereduce library in Python, which applies a non-stationary noise reduction algorithm with maximum noise suppression (prop_decrease = 1.0). The non-stationary setting adapts the noise profile across the duration of the recording, accounting for the fact that the background noise characteristics may vary as the vaporization event progresses. This aggressive denoising removes ambient laboratory noise while preserving the impulsive acoustic signatures associated with bubble nucleation and collapse events. The denoised signal is then trimmed using librosa.effects.trim with a threshold of 12 dB below the peak amplitude, which removes the portions of the recording captured before the droplet contacts the surface and after complete evaporation, isolating only the active vaporization event. The trimmed audio signal is then segmented into four equal-length temporal segments (quartiles), dividing the vaporization event into four consecutive time windows that correspond approximately to the initial spreading phase (Q1), peak nucleation activity (Q2), declining nucleation (Q3), and final evaporation (Q4). For each quartile segment, a Short-Time Fourier Transform (STFT) is computed using a window length of 128 samples (n_fft = 128) and a hop length of 32 samples, yielding a magnitude spectrogram with a frequency resolution of 375 Hz per bin (48,000 Hz / 128) and 65 frequency bins spanning from 0 to the Nyquist frequency of 24 kHz. Three features are then extracted from each quartile's STFT magnitude spectrogram:

Mean amplitude (a_1, a_2, a_3, a_4): The mean value of the STFT magnitude spectrogram across all frequency bins and time frames within the quartile. Although labeled "mean amplitude" for consistency with the associated conference publication [6], this feature is computed from the spectral domain rather than the raw time-domain waveform. It captures the overall acoustic energy in each temporal segment of the vaporization process, reflecting the cumulative intensity of all acoustic sources — bubble nucleation, growth, coalescence, departure, and collapse — active during that portion of the event.

Mean peak frequency (f_1, f_2, f_3, f_4): For each time frame in the STFT, the frequency bin with the maximum magnitude is identified as the dominant frequency for that frame. The mean peak frequency is computed as the average of these per-frame dominant frequencies across all time frames within the quartile. This feature captures the characteristic frequency of the most energetic acoustic events in each temporal segment. Through the Minnaert frequency relationship ($f \propto R^{-1}$), the dominant acoustic frequency correlates inversely with the characteristic bubble radius, providing an indirect measure of bubble size that evolves across the vaporization event.

Peak frequency standard deviation (σ_1, σ_2, σ_3, σ_4): The standard deviation of the per-frame dominant frequencies across all time frames within the quartile. This feature captures the variability of the dominant acoustic frequency over time — a narrow distribution (low σ) indicates relatively uniform bubble activity with consistent bubble sizes, while a broad distribution (high σ) indicates diverse bubble dynamics with a wide range of bubble sizes and multiple types of acoustic events occurring within the quartile.

This extraction yields 12 acoustic feature values per test case (a_1, a_2, a_3, a_4, f_1, f_2, f_3, f_4, σ_1, σ_2, σ_3, σ_4), representing a dramatic dimensionality reduction from the raw waveform (which contains tens of thousands of samples at 48 kHz) to a compact feature vector that preserves the essential statistical characteristics of the acoustic signal. In conjunction with the droplet volume, these 13 input features (12 acoustic plus droplet volume) form the input to the neural network architectures described in the following sections.

The quartile-based representation offers several advantages. It dramatically reduces the dimensionality of the acoustic input data while preserving the essential physical information. The quartile features are inherently robust to outliers and noise because they are computed as statistical summaries over large numbers of STFT frames. The temporal segmentation into quartiles preserves information about how the acoustic characteristics evolve during the vaporization process. As shown in Table 4.2 and Figure 4.2, this temporal evolution differs qualitatively between regimes: in the conduction regime, the amplitude decreases monotonically from Q1 through Q4 as the gentle evaporation gradually subsides, while in the nucleation regimes, the amplitude peaks sharply in Q2 as nucleation intensifies before decaying through Q3 and Q4 as the liquid volume diminishes. These regime-specific temporal patterns, encoded naturally by the quartile structure, provide the neural network with discriminating features beyond what a single whole-signal summary could capture.

4.3.2. How the Neural Network Distinguishes Between Regimes

The quantitative acoustic characterization reveals that the 12 quartile features provide multiple independent discriminating dimensions that enable the neural network to distinguish between the four boiling regimes. Three primary feature patterns serve as regime discriminators.

First, the Q2 amplitude provides the strongest single-feature separation between regimes. The conduction regime clusters tightly at Q2 values below 0.013, isolated bubble ranges from 0.009 to 0.083, vigorous nucleation clusters between 0.045 and 0.131 with low variability (standard deviation 0.017), and PDVR spans a broad range from 0.009 to 0.141 with high variability (standard deviation 0.039). The 14× difference in mean Q2 amplitude between conduction and vigorous nucleation provides a clear separation for the neural network, while the high variability in PDVR creates a distinguishing signature for that regime.

Second, the temporal amplitude pattern across the four quartiles encodes regime-specific dynamics that are clearly distinct. The conduction regime is the only regime exhibiting a monotonically decreasing amplitude profile (Q1 > Q2 > Q3 > Q4), reflecting steady evaporation without nucleation onset. In contrast, all three nucleation regimes exhibit Q2 peaks of increasing sharpness: isolated bubble at 2.7× the Q1 value, vigorous nucleation at 21×, and PDVR at 37×. This temporal shape information, encoded in the four amplitude features (a_1–a_4), provides the neural network with a qualitative regime signature that complements the quantitative amplitude magnitude. The conduction regime's declining profile is unambiguously separable from the peaked profiles of the nucleation regimes, while the steepness of the Q2 peak discriminates among the three nucleation regimes.

Third, the Q2 frequency separates regimes into two groups: conduction and isolated bubble below 3,500 Hz, and vigorous nucleation and PDVR above 4,500 Hz, as shown in
Figure 4.2 (b). Within each group, the amplitude features provide further discrimination. The frequency standard deviation features (σ_1–σ_4), shown in
Figure 4.2 (c), add additional discriminating power by encoding the breadth of the acoustic spectrum — isolated bubble exhibits the highest mean frequency standard deviation (2,009 Hz), while conduction exhibits the lowest (1,657 Hz).

4.3.3. The Continuous Spectrum of Temperature and Acoustic Response

The relationship between superheat temperature and acoustic features is continuous rather than categorical. Although the four regime labels provide a useful classification framework, the acoustic emissions evolve gradually with increasing superheat, reflecting the continuous nature of the underlying physics. The mean amplitude increases from 0.0034 at 10°C to approximately 0.0405 at 35–40°C, where it plateaus briefly, then decreases through the PDVR regime to 0.0243 at 55°C as progressive dryout reduces the wetted surface area available for bubble nucleation. The maximum amplitude, however, continues increasing from 0.0094 at 10°C to 0.1412 at 50°C, reflecting the increasing violence of individual vapor recoil events even as the overall number of active nucleation sites decreases.

This continuous variation is significant for the neural network architecture, which treats superheat prediction as a regression task and regime classification as a categorical task. The regression output can capture the smooth, continuous evolution of heat transfer performance with temperature, while the classification output captures the discrete transitions between qualitatively different physical regimes.

The acoustic features support both tasks simultaneously because they encode both the continuous intensity trend (through amplitude features) and the regime-specific temporal and spectral patterns (through the quartile structure and frequency features). The fact that the multimodal fusion network achieves 4.6% RMSPE in heat flux regression and 100% regime classification accuracy demonstrates that the network successfully exploits both the continuous and categorical information content in the acoustic data.

4.4. Acoustic-Only Neural Network

The first neural network architecture developed in this study was designed to evaluate the standalone predictive capability of acoustic data, isolating the information content of acoustic emissions from the other sensing modalities. This network, depicted in Figure 4.4, takes the 12 quartile features along with the initial droplet volume V_0 as inputs (13 total inputs) and predicts the surface superheat temperature.

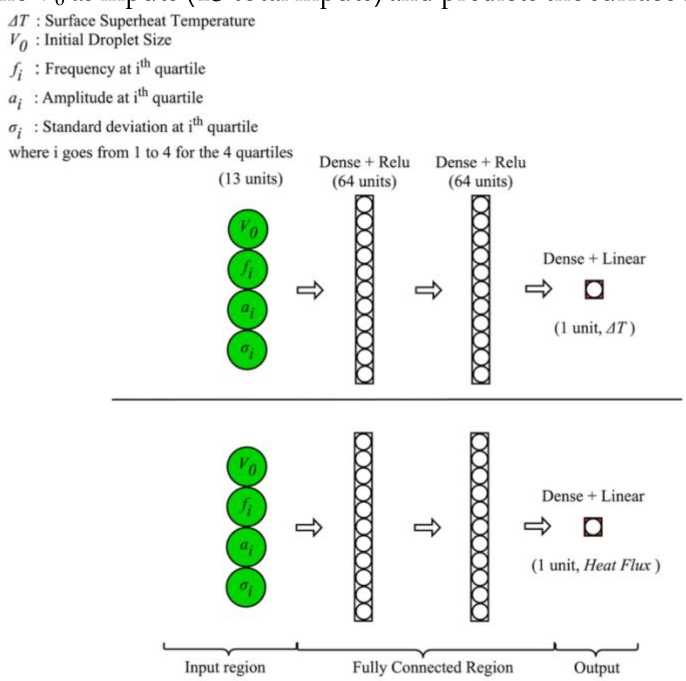


Figure 4.4. Two Neural network architectures for predicting superheat temperature and Heat Flux from acoustic quartile features.

The acoustic-only network consists of two fully-connected dense layers with 64 neurons each. The input layer accepts the 13 features (12 acoustic quartile features plus droplet volume). The first hidden layer contains 64 neurons with ReLU activation, followed by a second hidden layer with 64 neurons and ReLU activation. The output is a single-neuron layer predicting the superheat temperature. This same architecture was replicated to train a separate model for heat flux prediction. The network was trained using the Adam optimizer with a learning rate of 0.001 and mean absolute error loss function.
The results of the acoustic-only neural network are presented in Figure 4.5. The acoustic-only model achieved a RMSPE of 13.3% for superheat prediction and 32.9% for heat flux prediction on the unseen test data. While the heat flux prediction error is too large for practical applications on its own, the superheat prediction accuracy of 13.3% is remarkable given that the input consists solely of statistical

summaries of acoustic signals. This demonstrates that acoustic emissions during droplet vaporization contain sufficient information to estimate the thermal state of the boiling system within a reasonable accuracy.

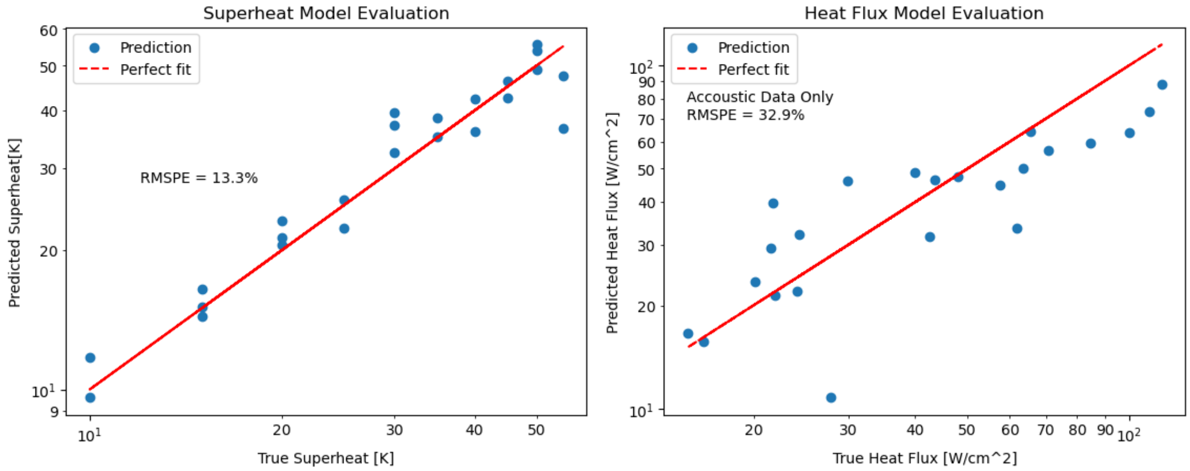


Figure 4.5. Acoustic-only prediction: (a) superheat prediction, RMSPE = 13.3%; (b) heat flux prediction, RMSPE = 32.9%.

Analysis of the prediction errors reveals that the acoustic model performs best in the intermediate superheat range (20–45°C), corresponding to the isolated bubble and vigorous nucleation regimes. The largest errors occur at low superheats (where acoustic emissions are minimal and close to the noise floor) and at very high superheats in the PDVR regime (where the chaotic nature of the process introduces significant acoustic variability between trials at the same nominal superheat). The physics underlying the acoustic-superheat correlation can be understood through the fundamental relationships between bubble dynamics and sound generation: higher superheat temperatures lead to more vigorous nucleation, larger bubble departure diameters, and more frequent bubble collapse events, all of which produce more intense acoustic emissions. The quartile-based feature extraction effectively captures this relationship; the mean absolute amplitude per quartile increases systematically with increasing superheat temperature.

4.5. Comparison with Image-Only Prediction

To contextualize the acoustic-only prediction results, we compare them with the Case B image-only prediction results from Chapter 3. Both approaches represent non-thermal sensing modalities that attempt to infer thermal state from indirect measurements. The comparison, shown in Table 4.3, reveals that the image-based Case B model outperforms the acoustic-only model for both superheat prediction (10.9% vs. 13.3% RMSPE) and heat flux prediction (14% RMSPE on image only prediction cascaded to Case C vs. 32.9% RMSPE on acoustic).

Table 4.3. Comparison Between Image-Only and Acoustic-Only Prediction

Input Modalities	Superheat (Test)	Heat Flux RMSPE (Test)
Image Only	10.9%	14% (Cascade to Case C)

Acoustic Only	13.3%	32.9%

This difference is expected given that high-speed video imagery captures detailed spatial information about bubble morphology, liquid film geometry, and nucleation site distribution that directly reflects the thermal state of the system. In contrast, the acoustic signal represents a temporally-integrated measure of bubble activity that lacks spatial resolution, the microphone captures a superposition of all acoustic events occurring on the surface without information about their spatial locations.

However, the acoustic modality offers distinct practical advantages that complement image-based sensing. Acoustic sensors are significantly less expensive than high-speed cameras, require no optical access to the boiling surface, are immune to optical obscuration from steam or liquid splashing, and can operate in environments where imaging is impractical (such as inside pressure vessels or in geometrically constrained configurations). The 13.3% RMSPE achieved by acoustic sensing, while less precise than imaging, may be sufficient for many monitoring and control applications. Furthermore, as demonstrated in the following section, the combination of acoustic and image data yields improved predictions compared to either modality alone, demonstrating that acoustic sensing captures complementary information not fully represented in visual data.

4.6. Multimodal Fusion Network

Building on the hybrid CNN architecture from Chapter 3, we developed an extended multimodal fusion network that integrates acoustic data with image and digital sensor inputs. The multimodal network, depicted in

Figure 4.6, employs a parallel-series hybrid architecture with three input branches.

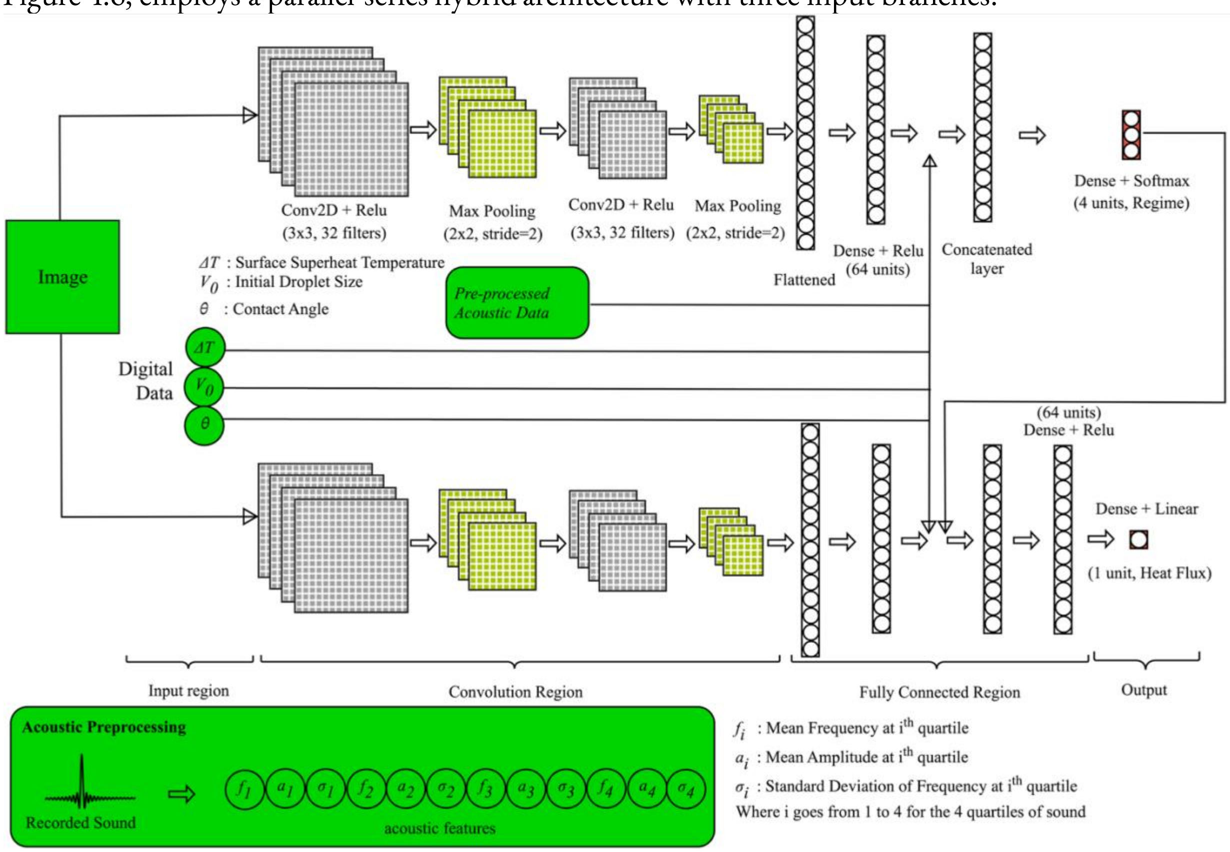


Figure 4.6. Multimodal fusion network architecture incorporating image, acoustic, and digital sensor branches.

The hybrid parallel-series architecture processes the input image through two independent convolutional branches, each consisting of two convolutional layers (32 filters each, 3×3 kernels) with ReLU activation and 2×2 max pooling, followed by flattening and a Dense layer with 64 neurons. Although both branches receive the same image input, they learn different feature representations because they are trained by different loss gradients — the first branch is optimized for regime classification while the second is optimized for heat flux prediction. The digital sensor data (superheat temperature, droplet volume, contact angle) and all 12 acoustic quartile features (a_1–a_4, f_1–f_4, σ_1–σ_4) are provided as skip connection inputs to both output paths.

The "series" component of the architecture connects the two tasks sequentially: the regime classification output from the first branch — a four-element softmax probability vector — is concatenated with the second convolutional branch output, the digital sensor data, and the acoustic features before passing through a Dense layer with 64 neurons and ReLU activation to produce the heat flux prediction. This cascaded design allows the predicted regime probabilities to inform the heat flux prediction, reflecting the physical reality that the heat transfer mechanism differs qualitatively across regimes and the regime context helps the network select the appropriate mapping from input features to heat flux.

The multimodal fusion network was trained using the same optimizer (Adam, learning rate 0.001), loss function (mean absolute error for regression, categorical crossentropy for classification), and data split (85% training, 15% testing) as the previous architectures. The image data augmentation procedure was applied identically to the procedure described in Section 3.2.1.

4.7. Results and Discussion

The multimodal fusion network achieved a mean heat flux prediction RMSPE of 1.1% on the training dataset and 4.6% on the testing dataset, as shown in Figure 4.7. This represents a 20.7% relative reduction in testing error compared to the Case C baseline from Chapter 3 (from 5.8% to 4.6% RMSPE). The improvement confirms that acoustic data provides complementary information that enhances prediction accuracy beyond what can be achieved with visual and thermal data alone.

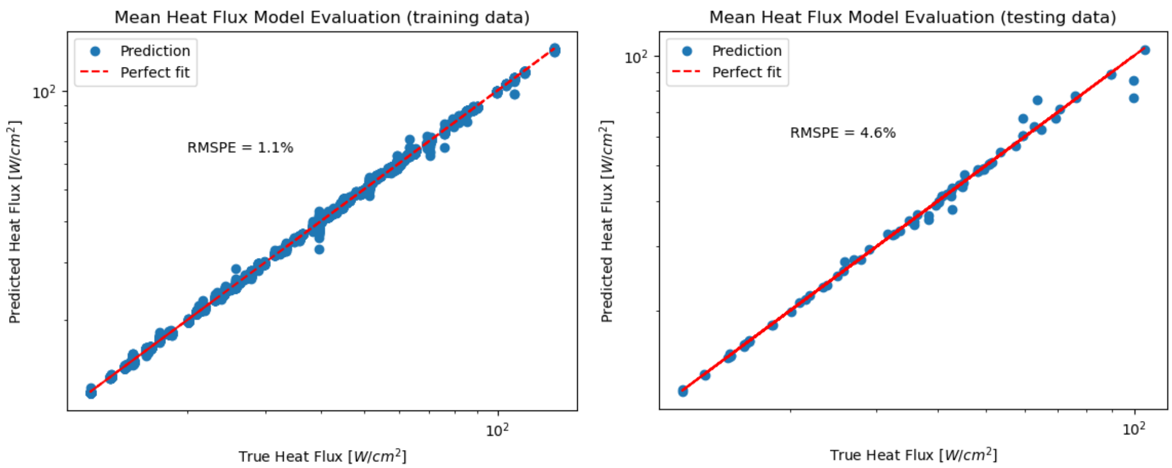


Figure 4.7. Mean heat flux prediction results for the multimodal fusion network on (a) the training and (b) the testing datasets.

For regime classification, the multimodal fusion network achieved 100% accuracy on both the training and testing datasets (80 test images), as depicted in Figure 4.8. This is a meaningful improvement over the Case C model, which misclassified 2 out of 65 images (96.9% accuracy). The two images misclassified by Case C occurred at regime boundaries where the visual morphology is ambiguous. The addition of acoustic data resolves this ambiguity because the acoustic signatures of adjacent regimes are distinct even when the visual appearance is transitional. For example, the onset of PDVR produces characteristic high-amplitude acoustic bursts from vapor recoil events that are acoustically distinguishable from vigorous nucleation even when the visual morphology appears similar.

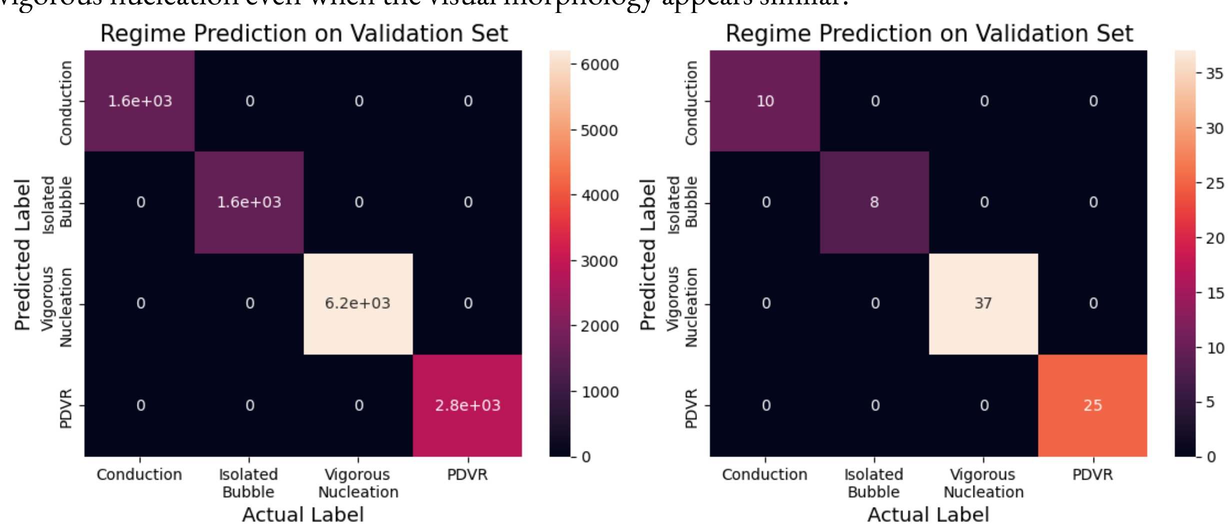


Figure 4.8. Nucleation regime prediction for the multimodal fusion network: (a) training dataset, (b) testing dataset.

The improvement in heat flux prediction is most pronounced in the vigorous nucleation and PDVR regimes, where the complex two-phase morphology benefits most from the additional information provided by acoustic sensing. In these regimes, visual observation can be complicated by rapid bubble dynamics and chaotic liquid ejection, while the acoustic signatures of these events remain clearly detectable and provide additional discriminative information for the neural network.

The comprehensive comparison across all architectures demonstrates the progressive improvement: Case A (digital only) 15.8%, Case B (image + non-thermal) 10.3%, Case C (image + digital) 5.8%, and multimodal fusion (image + digital + acoustic) 4.6% test RMSPE. This comparison is summarized in Table 4.4.

Table 4.4. Comparison Between Multimodal Fusion and Previous Architectures

Architecture	Input Modalities	Heat Flux RMSPE (Train)	Heat Flux RMSPE (Test)
Case A	Digital Only	12.9%	15.8%
Case B	Image + non-thermal	2.0%	10.3%
Case C	Image + all digital	1%	5.8%
Multimodal Fusion	Image + digital + acoustic	1.1%	4.6%

4.8. Feature Importance Analysis

To understand the relative contribution of each input modality, a permutation importance analysis was applied. This technique evaluates the importance of each input feature by measuring the degradation in prediction accuracy when that feature's values are randomly shuffled (permuted) across the dataset, breaking the relationship between the feature and the output while preserving the marginal distribution of the feature values. A feature whose permutation causes a large increase in prediction error is important for the model's predictions, while a feature whose permutation has little effect is relatively unimportant. The permutation importance results are presented in Figure 4.9. The analysis reveals several important findings. First, superheat temperature is not among the most important input features for either heat flux prediction or regime classification. This may seem counterintuitive given that superheat is the primary independent variable in boiling experiments, but it can be understood by recognizing that the image features already encode the thermal state implicitly through the visual morphology (as demonstrated by Case B's ability to predict superheat from images and non-thermal data), and the acoustic features encode it through the intensity and frequency of nucleation events.

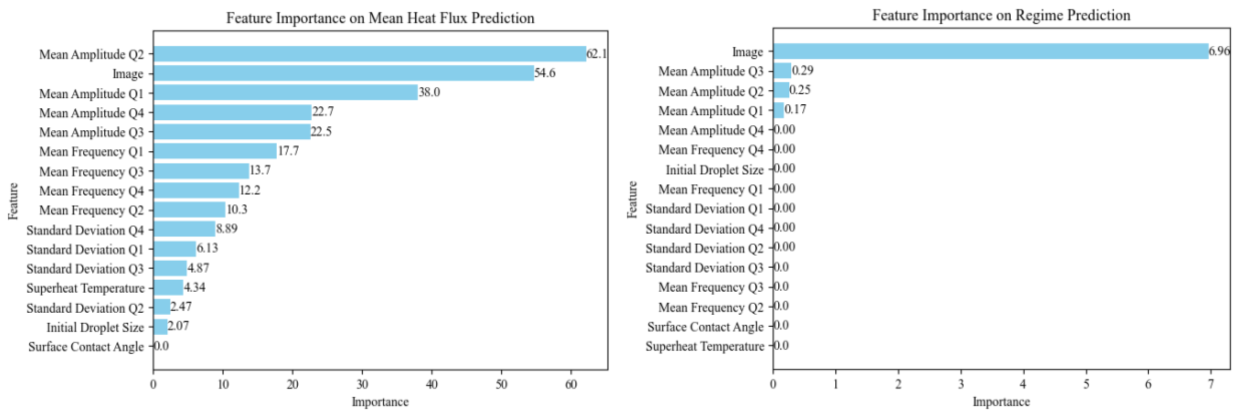


Figure 4.9. Permutation feature importance showing the relative contribution of each input feature to the prediction accuracy for (a) mean heat flux prediction and (b) regime classification.

Second, image and acoustic features contribute most to heat flux predictions. The acoustic quartile features collectively contribute significantly to prediction accuracy, with the mean amplitude features (a_1–a_4) being particularly important because they directly reflect the overall intensity of bubble activity. Third, the image features show moderate importance for heat flux prediction but are primarily important for regime classification, where the spatial distribution and morphology of bubbles provide the most discriminative information.

This confirms that each modality captures complementary information about the boiling process: images capture spatial morphology (bubble sizes, wetting areas, dryout patterns), acoustic data captures temporal dynamics (nucleation frequency, collapse intensity, vapor recoil events), and digital sensors capture boundary conditions (temperature, volume, contact angle) that constrain the physical system.

4.9. Physical Interpretation of Acoustic Features

The quartile-based acoustic feature extraction developed in this chapter provides not only an effective input representation for the neural network but also physically interpretable insights into the boiling process. Analysis of the acoustic feature values across the superheat range reveals several systematic trends that connect directly to the underlying bubble dynamics.

The mean amplitude (a_1–a_4) increases with superheat temperature from the conduction through the vigorous nucleation regime, reaching a maximum at approximately 35–40°C superheat, then decreases in the PDVR regime as progressive dryout reduces the wetted surface area. The maximum amplitude, however, continues to increase through the PDVR regime, reaching its highest value of 0.1412 at 50°C superheat, reflecting the increasing violence of individual vapor recoil events even as the overall mean acoustic intensity declines. This trend reflects the increasing intensity of acoustic sources: at low superheats, the primary acoustic source is the gentle evaporation at the liquid-air interface, producing minimal acoustic emissions. As superheat increases, bubble nucleation events generate discrete pressure pulses that increase in both amplitude and frequency of occurrence. In the vigorous nucleation regime, the superposition of many simultaneous bubble events creates a high-amplitude broadband signal. In the PDVR regime, the amplitude becomes highly variable, with intermittent very-high-amplitude bursts from vapor recoil ejection events separated by periods of lower amplitude when less of the surface is in liquid contact.

The temporal evolution of amplitude across the four quartiles (Q1 through Q4) provides information about the dynamics of the vaporization event. For the conduction regime, the amplitude decreases monotonically from Q1 through Q4, reflecting the steady, gradually subsiding nature of the evaporation process without discrete nucleation events. For the nucleate boiling regimes, the amplitude typically increases from Q1 to Q2 or Q3 as the nucleation process intensifies, then decreases in Q4 as the liquid volume diminishes and fewer nucleation sites remain active. For the vigorous nucleation regime, the amplitude typically increases from Q1 to Q2 with a steeper slope than its decrease in Q3 and Q4 as the liquid volume diminishes. For the PDVR regime, the amplitude pattern is more variable, reflecting the stochastic nature of vapor recoil events that can occur at any point during the vaporization process.

The mean frequency features (f_1–f_4) capture the characteristic frequency of the dominant acoustic sources. The natural frequency of bubble oscillation depends on bubble radius as approximately $f \propto R^{-1}$ (the Minnaert frequency relationship), so the dominant acoustic frequency decreases as bubble sizes increase at higher superheats. However, the observed frequency trend is more complex because multiple bubble sizes coexist and the acoustic spectrum is broadband. The frequency standard deviation features (σ_1–σ_4) capture the breadth of the frequency distribution, with broader distributions indicating a wider range of bubble sizes and dynamics. The vigorous nucleation and PDVR regimes typically exhibit the broadest frequency distributions, reflecting the diversity of bubble sizes and the multiple types of acoustic events (nucleation, coalescence, departure, vapor recoil) occurring simultaneously.

These physical interpretations of the acoustic features confirm that the quartile-based extraction captures meaningful physical information about the boiling process rather than arbitrary statistical artifacts. The neural network's ability to use these features for accurate superheat and heat flux prediction further validates that the quartile representation preserves the essential acoustic information content despite the dramatic dimensionality reduction from the raw waveform.

4.10. Computational Efficiency

An important practical consideration for any machine learning approach intended for real-time monitoring applications is computational efficiency. The multimodal fusion network developed in this chapter is designed to be lightweight enough for real-time deployment while achieving high prediction accuracy. The total number of trainable parameters in the multimodal fusion network is approximately 1.3 million. The inference time for a single prediction (one image plus corresponding digital and acoustic features) is on the order of milliseconds on modern hardware, enabling real-time processing at rates exceeding 100 predictions per second, far faster than the camera frame rate of 1200 fps when applied to individual frames, and fast enough for the per-event analysis used in this study where each vaporization event produces a single prediction.

The quartile-based acoustic feature extraction is particularly efficient because it reduces the raw acoustic waveform (tens of thousands of samples at 48 kHz) to just 12 features through simple statistical operations (mean, standard deviation) that can be computed in microseconds. This is significantly more efficient than spectrogram-based approaches, which require computing Short-Time Fourier Transforms over multiple overlapping windows and produce two-dimensional feature maps that require additional convolutional processing. The efficiency of the quartile approach makes it well-suited for edge deployment on embedded systems with limited computational resources, as will be discussed in Section 6.6. The next chapter discusses how multimodal information analysis can enhance boiling research.

4.11. Conclusion

This chapter has demonstrated the value of acoustic emission monitoring as a complementary sensing modality for boiling heat transfer analysis. The key findings are:

First, acoustic signals generated during water droplet vaporization on superheated nanostructured surfaces contain information that correlates with surface superheat temperature and boiling regime. A neural network trained on quartile-based acoustic features achieved 13.3% RMSPE in predicting superheat temperatures ranging from 10°C to 55°C, demonstrating that acoustic data captures latent physical information intrinsic to the boiling process.

Second, multimodal fusion achieves 4.6% RMSPE with 20.7% improvement over the previous best Case C result, and 100% regime classification accuracy on the test dataset, confirming that acoustic data provides complementary information beyond what image and digital data alone can capture.

Third, the quartile-based acoustic feature extraction provides an efficient representation that captures essential signal characteristics while dramatically reducing data dimensionality from tens of thousands of raw waveform samples to just 12 summary features per test case.

Fourth, acoustic sensing is particularly valuable in the PDVR regime (40–55°C superheat), where visual observation is complicated by chaotic liquid ejection due to vapor recoil forces. The violent acoustic signatures of vapor recoil events remain clearly detectable and provide information not readily available from visual data.

Fifth, the multimodal sensor fusion strategy (combining visual, thermal, and acoustic sensing) represents a promising approach for analyzing complex boiling systems that is poised to significantly improve model accuracy and robustness for analyzing physical systems that are best investigated experimentally.

Chapter 5.
Assessment of Multimodal Information Analysis Enhancement Boiling Research

Chapters 3 and 4 developed and evaluated a series of CNN architectures that progressively incorporated additional sensing modalities — from digital-only to image-plus-digital to full multimodal fusion with acoustic emission data. Each chapter focused on the architecture development and performance evaluation for a specific modality configuration. This chapter steps back from individual architectures to assess the multimodal approach as a whole, examining the trajectory of improvement across all configurations, analyzing what each modality contributes physically, identifying the mechanisms underlying the multimodal advantage, and evaluating how close the achieved accuracy is to the fundamental limits imposed by experimental uncertainty.

5.1. Introduction

The preceding chapters have presented a systematic progression from single-modality digital data analysis through dual-modality image-digital fusion to trimodal fusion incorporating acoustic emission data. Each step in this progression yielded measurable improvement in both heat flux prediction accuracy and regime classification performance, culminating in the multimodal fusion network's 4.6% RMSPE and 100% regime classification accuracy reported in Chapter 4. However, these results were presented within the context of their respective chapters, making it difficult to assess the overall trajectory and draw cross-cutting conclusions about the value of multimodal sensing for boiling heat transfer research.

This chapter synthesizes the findings across all architectures to provide a comprehensive assessment of how multimodal information analysis enhances boiling heat transfer research. The assessment addresses four key questions: (1) What is the quantitative improvement trajectory as modalities are added? (2) What complementary information does each modality contribute, and what are the physical mechanisms underlying the multimodal advantage? (3) How close is the achieved accuracy to the fundamental limits imposed by experimental uncertainty and the inherent stochasticity of the boiling process? (4) What are the practical implications for experimental boiling research and thermal system monitoring?

5.2. Progressive Improvement Trajectory

The most striking finding of this dissertation is the consistent, monotonic improvement in prediction accuracy as additional sensing modalities are incorporated into the analysis framework. Table 4.4 consolidates the key performance metrics across all architectures developed in Chapters 3 and 4, presented in order of increasing modality count. The trajectory from 15.8% to 4.6% RMSPE represents a 70.9% total relative reduction in prediction error achieved through the systematic incorporation of additional sensing modalities. This improvement can be decomposed into two major steps. The first step, from Case A to Case C, reflects the addition of image data to digital sensor measurements. This step reduces the error from 15.8% to 5.8%, a 63.3% relative improvement, demonstrating that image data contains substantial quantitative information about the boiling process that is not captured by digital measurements alone. The second step, from Case C to multimodal fusion, reflects the addition of acoustic data. This step reduces the error from 5.8% to 4.6%, a 20.7% relative improvement, demonstrating that acoustic emissions contain complementary information beyond what imaging and digital sensors capture.

The diminishing marginal improvement, 63.3% from adding images versus 20.7% from adding acoustics, does not indicate that acoustic data is less valuable than image data. Rather, it reflects the fundamental principle that the first additional modality captures the largest pool of previously unmeasured information, while subsequent modalities capture progressively smaller (but still significant) pools of complementary information. This pattern is consistent with the information-theoretic prediction that the mutual information between each new modality and the prediction target, conditioned on the existing modalities, decreases as more modalities are included. The important finding is that the improvement from acoustic data is substantial (20.7%) and statistically meaningful, confirming that acoustic emissions carry unique information about the boiling process that cannot be inferred from visual or digital data alone.

The regime classification results tell a similar but even more striking story. Case C achieved 96.9% accuracy (misclassifying 2 of 65 test images), while the multimodal fusion network achieved 100% accuracy on 80 test images. The two images misclassified by Case C both occurred at the boundary between the vigorous nucleation and PDVR regimes, where the visual morphology is most ambiguous. The addition of acoustic data resolves this ambiguity because the acoustic signatures of these two regimes are distinct, PDVR produces characteristic high-amplitude bursts from vapor recoil events that are acoustically distinguishable from the continuous acoustic emission of vigorous nucleation, even when the visual appearance is transitional.

5.3. Complementary Information Content of Each Modality

The permutation importance analysis presented in Section 4.8 provides quantitative evidence for the complementary nature of information captured by each modality, showing that image features, acoustic features, and digital sensor data each contribute independently to prediction accuracy and cannot substitute for one another. This section provides a deeper physical interpretation of these findings by examining what each modality captures about the boiling process and why these contributions are fundamentally non-redundant. The analysis is organized by modality, examining digital sensor data, image data, and acoustic data in turn, before discussing how their combination enables the multimodal advantage observed in the prediction results.

5.3.1. Digital Sensor Data: Boundary Conditions and State Variables

Digital sensor data, superheat temperature, droplet volume, and contact angle, capture the boundary conditions and thermodynamic state variables that constrain the physical system. The superheat temperature determines the driving force for heat transfer and defines the thermodynamic potential for bubble nucleation. The droplet volume determines the total thermal energy required for complete vaporization and constrains the maximum wetted area. The contact angle characterizes the surface wettability and affects the initial spreading dynamics and the capillary forces that supply liquid to nucleation sites.

The fundamental limitation of digital-only models (Case A, 15.8% RMSPE) is that these boundary conditions do not fully determine the instantaneous state of the boiling process. For a given set of boundary conditions, the stochastic nature of bubble nucleation means that the actual two-phase morphology, and consequently the instantaneous heat transfer performance, can vary significantly from one experimental realization to another. Digital sensors capture the average conditions but cannot capture the specific morphological state at each measurement instant.

5.3.2. Image Data: Spatial Morphology and Two-Phase Structure

High-speed video imagery captures the spatial morphology of the two-phase mixture on the heated surface. The visual features that the CNN learns to extract include bubble sizes and shapes, which reflect the balance between surface tension, buoyancy, and vapor recoil forces at each nucleation site. The CNN also learns from the spatial distribution of nucleation sites, which determines the fraction of the surface involved in active nucleate boiling versus conduction-dominated evaporation. The extent and geometry of wetted versus dry regions, which directly control the local heat transfer coefficient, provide critical spatial information. The dynamics of liquid film evolution, including spreading, recession, and breakup, are captured across sequential video frames.

The substantial improvement from Case A to Case B (15.8% to 10.3% RMSPE, a 34.8% relative reduction) demonstrates that these spatial features contain significant predictive power even without direct temperature information. The further improvement from Case B to Case C (10.3% to 5.8% RMSPE, a 43.7% relative reduction) demonstrates that the combination of spatial features and temperature information is more powerful than either alone, the image features provide context that helps the network interpret the temperature data more effectively (and vice versa), enabling more accurate predictions than simply adding the individual contributions.

The ability of Case B to predict superheat temperature from images alone (7.0% RMSPE) confirms that the visual morphology encodes the thermal state of the system through the well-known correlation between superheat level and boiling regime. The CNN learns this correlation from the data without being explicitly programmed with boiling physics, demonstrating the power of convolutional feature extraction for capturing physically meaningful patterns in experimental images.

5.3.3. Acoustic Data: Temporal Dynamics and Subsurface Events

Acoustic emissions capture temporal dynamics of the boiling process at timescales and in locations that are inaccessible to imaging. The physical mechanisms generating acoustic emissions during boiling include bubble nucleation (the rapid expansion of a vapor bubble creates a pressure pulse in the surrounding liquid), bubble growth oscillation (growing bubbles can oscillate at their natural frequency, generating continuous acoustic emission), bubble departure and detachment (the separation of a bubble from the surface creates an acoustic impulse), bubble collapse and condensation (bubbles that enter subcooled liquid regions collapse, generating intense pressure pulses), and vapor recoil events in the PDVR regime (the violent ejection of liquid by high-velocity vapor produces distinctive acoustic signatures).

These acoustic events occur at temporal frequencies up to several kHz, well above the camera frame rate of 1,200 fps. This means that acoustic emissions can capture multiple bubble nucleation, growth, and collapse cycles between consecutive video frames, providing temporal information about the boiling process that is fundamentally inaccessible through imaging, regardless of camera quality. The acoustic sensor also captures events occurring beneath the liquid surface or in regions obscured from the camera's view by steam, bubbles, or ejected liquid, providing information about subsurface bubble dynamics that is invisible to top-down imaging.

The 13.3% RMSPE achieved by acoustic-only superheat prediction demonstrates that these temporal dynamics contain substantial information about the thermal state of the system. The much larger error for acoustic-only heat flux prediction (32.9%) reflects the fact that heat flux depends strongly on the spatial distribution of nucleation activity and the wetted area fraction, information that is better captured by imaging than by acoustic sensing alone.

The 20.7% relative improvement from adding acoustic data to the image-plus-digital baseline (from 5.8% to 4.6% RMSPE) confirms that acoustic information is genuinely complementary to visual and digital data. The improvement is not simply a matter of having more data, it reflects the fact that acoustic emissions capture physical phenomena (temporal dynamics, subsurface events, high-frequency oscillations) that are not represented in the other modalities.

5.4. Physical Mechanisms of Multimodal Advantage

The physical basis for the multimodal advantage can be understood by examining how different modalities capture different aspects of the boiling process across the regime spectrum. In the conduction-dominated regime, digital sensors provide the most informative data because the process is quasi-steady and well-described by the boundary conditions. Imaging provides limited additional information because the visual morphology (a thin, spread liquid film with minimal bubble activity) is relatively uniform across the superheat range within this regime. Acoustic emissions are minimal, contributing little to prediction accuracy. The multimodal advantage in this regime is correspondingly small.

In the isolated bubble and vigorous nucleation regimes, the multimodal advantage becomes significant. Imaging captures the spatial distribution of nucleation sites, bubble sizes, and the evolving liquid film geometry, providing substantial information beyond what digital sensors alone can measure. Acoustic emissions capture the frequency of nucleation events, the intensity of bubble collapse, and the temporal pattern of bubble activity, adding a temporal dimension to the spatial information from imaging. The combination enables the network to build a more complete representation of the boiling process, accounting for both spatial and temporal aspects of nucleation dynamics.

In the PDVR regime, the multimodal advantage is greatest. Visual observation becomes challenging due to the chaotic ejection of liquid and rapid changes in morphology from vapor recoil forces. The camera captures instantaneous snapshots of a rapidly evolving process, but the morphology can change dramatically between frames at 1,200 fps. Acoustic emissions, in contrast, capture the dynamics of vapor recoil events as they occur (at up to 24 kHz Nyquist frequency), providing a continuous record of the intensity and frequency of these violent events. The digital sensors capture the average thermal conditions but cannot distinguish between different states of the surface dryout cycle. The fusion of all three modalities provides the most robust characterization of this challenging regime, which is why the improvement from adding acoustic data is most pronounced in the PDVR regime.

5.5. Approaching Fundamental Accuracy Limits

The 4.6% RMSPE achieved by the multimodal fusion network raises the question of how close this result is to the fundamental limits of prediction accuracy for this system. Several sources of uncertainty exist in the experimental system. The thermocouple measurement uncertainty of ±2°C introduces uncertainty in the reference superheat values used as both input and for evaluating prediction accuracy. The droplet volume uncertainty of ±0.3 μL (from pipette calibration) on droplets of 5–8 μL contributes approximately ±4–6% relative uncertainty to the heat flux calculation through V_0 in Equation 3.1. The evaporation time uncertainty of approximately ±2 frames at 1200 fps contributes less than ±1% for typical evaporation events spanning tens to hundreds of frames. Additionally, the average wetted area is approximated as one-half the maximum wetted area (A_avg = A_max/2), which assumes a linear decrease in wetted area over the course of vaporization. The actual wetted area profile may deviate from linearity, particularly in the PDVR regime where chaotic liquid ejection creates non-monotonic changes in wetted area, introducing a systematic approximation error that has not been independently quantified in this study. The droplet volume uncertainty alone places a lower bound of approximately ±4–6% on the heat flux uncertainty, and the additional contributions from wetted area measurement and the A_max/2 approximation are expected to increase this further. The 4.6% RMSPE achieved by the multimodal model is therefore comparable in magnitude to the quantifiable instrument uncertainty, suggesting that the model captures a substantial fraction of the deterministic information content in the combined sensor data. Precise determination of whether the model has reached the experimental accuracy floor would require independent quantification of the wetted area uncertainty through frame-by-frame area tracking, which is recommended as future work.

The 4.6% RMSPE achieved by the multimodal fusion network raises the question of how close this result is to the fundamental limits of prediction accuracy for this system. Several sources of irreducible uncertainty exist in the experimental system. The thermocouple measurement uncertainty of ±2°C introduces uncertainty in the reference superheat values used as both input and for evaluating prediction accuracy. The droplet volume uncertainty of ±0.3 μL affects both the heat flux calculation (through V_0 in Equation 3.1) and the prediction inputs. The contact angle uncertainty of ±0.3° introduces uncertainty in the surface characterization. The heat flux calculation itself has an uncertainty of approximately ±5% due to the approximation of the average wetted area as one-half the maximum wetted area and the assumption that the sensible heat contribution is negligible. The heat flux uncertainty can be estimated through standard error propagation applied to Equation 3.1. The mean heat flux is calculated as:

$$\overline{q"} = \frac{\rho V_0 h_{lv}}{At}$$

where ρ is the liquid water density, V_0 is the initial droplet volume, h_{lv} is the latent heat of vaporization, A is the average wetted area (approximated as one-half the maximum wetted area), and $t = \frac{n}{f}$ is the evaporation time determined from the frame rate of the high-speed video camera and the frame count between time of contact and complete evaporation. Since ρ, h_{lv}, and f are known material properties or instrument settings with negligible uncertainty, the relative uncertainty in heat flux is dominated by three terms:

$$\left(\frac{\delta\overline{q''}}{\overline{q''}}\right)^2 = \left(\frac{\delta V_0}{V_0}\right)^2 + \left(\frac{\delta A}{A}\right)^2 + \left(\frac{\delta t}{t}\right)^2$$

The droplet volume uncertainty ($\delta V_0 \approx \pm 0.3$ μL from pipette calibration on droplets of 5–8 μL) contributes approximately ±4–6% relative uncertainty. The evaporation time uncertainty ($\delta t \approx \pm 2$ frames at 1200 fps, from the subjective identification of the first-contact and complete-evaporation frames) contributes less than ±1% for typical evaporation events spanning tens to hundreds of frames. The average wetted area introduces the largest uncertainty through two sources: the measurement of the maximum wetted area from image analysis (estimated at ±3–5% from pixel-level boundary identification) and the systematic approximation that $A = A/2$, which assumes a linear decrease in wetted area from maximum to zero. If the actual wetted area profile deviates from linearity, this introduces a systematic bias that is difficult to quantify precisely but is estimated at ±5–10% based on comparison with frame-by-frame wetted area measurements for selected test cases. Combining these contributions in quadrature yields an overall heat flux uncertainty of approximately ±5–8%, depending on droplet size and evaporation conditions. This estimate provides important context for interpreting the multimodal fusion model's 4.6% RMSPE: because the heat flux labels used for training and evaluation are themselves uncertain by ±5–8%, the model's prediction error is approaching the floor set by experimental measurement precision. Further reduction in RMSPE would require improving the experimental measurements — for example, through automated frame-by-frame wetted area tracking to reduce the A approximation error, or through gravimetric droplet dispensing to reduce volume uncertainty — rather than through additional sensing modalities or more complex network architectures.

Beyond measurement uncertainty, the inherent stochasticity of the boiling process introduces irreducible variability. For identical nominal conditions (same superheat, droplet volume, and contact angle), the specific locations and timing of bubble nucleation events will vary from trial to trial due to the random nature of heterogeneous nucleation. This stochastic variability sets a fundamental floor on prediction accuracy that cannot be reduced through additional sensing modalities or more sophisticated machine learning architectures.

The 4.6% RMSPE approaches the estimated measurement uncertainty of approximately 5%, suggesting that the multimodal CNN captures nearly all of the deterministic information content in the combined measurement suite. Further improvement would likely require reducing the experimental uncertainty (e.g., through more precise temperature measurement, better control of droplet deposition conditions, or higher-resolution imaging) rather than adding additional sensing modalities or more complex network architectures.

5.6. Practical Implications

The findings of this dissertation have several practical implications for experimental boiling research and thermal system monitoring. For experimental research, the multimodal approach provides significantly more information about the boiling process from each experiment than traditional single-modality approaches. The ability to predict heat flux with 4.6% RMSPE from combined sensor data enables rapid parametric studies where the model can be used to estimate heat flux from easily measured quantities, reducing the need for labor-intensive post-processing of every experimental run. The regime classification capability (100% accuracy) provides automated regime identification that eliminates the subjective judgment inherent in manual classification from visual observation alone.

For thermal system monitoring and control, the demonstrated capability of acoustic-only superheat prediction (13.3% RMSPE) opens the possibility of non-intrusive boiling state monitoring using inexpensive acoustic sensors in environments where direct temperature measurement and optical access are impractical. In nuclear reactor cooling systems, for example, acoustic monitoring could provide early warning of transitions toward critical heat flux conditions without requiring modification of the reactor internals. In data center cooling systems, acoustic monitoring could enable real-time detection of boiling regime transitions in immersion-cooled electronics, triggering control actions to prevent thermal damage.

The quartile-based acoustic feature extraction developed in this dissertation provides a practical framework for real-time acoustic monitoring that is computationally efficient enough for deployment in embedded systems. Unlike spectrogram-based approaches that require computing full frequency transforms of long waveform segments, the quartile feature extraction reduces each acoustic signal to just 12 summary values that can be computed in real time and fed directly to a lightweight neural network predictor.

5.7. Comparison with Literature

The 4.6% RMSPE achieved by the multimodal fusion network compares favorably with the best prediction accuracy reported in the boiling heat transfer literature. Most neural network models for boiling heat transfer prediction report test errors in the range of 10–30% (e.g., Qiu et al. [10] for flow boiling, Calati et al. [11] for pool boiling in metal foams). The Dunlap et al. [62] acoustic-only study achieved 4.5% MAPE for pool boiling heat flux prediction using Gaussian Process Regression, but this was for a pool boiling configuration (constant heater power) rather than the more challenging droplet vaporization configuration with its transient dynamics and complex spreading behavior.

It is instructive to contextualize the multimodal CNN results against the accuracy of physics-based correlations that have historically been the primary predictive tools for boiling heat transfer. Standard nucleate boiling correlations such as those developed by Rohsenow [53], Cooper, and Stephan and Abdelsalam typically achieve prediction errors of 15–30% when applied broadly across different fluids, surfaces, and operating conditions, reflecting the challenge of capturing the complex, surface-dependent physics of nucleate boiling in a compact algebraic form. When these correlations are carefully fitted to high-quality data from a specific surface and fluid combination, prediction errors can be reduced to approximately 5–10%, representing the practical accuracy ceiling for physics-based approaches on well-characterized systems. The Case A digital-only neural network (15.8% RMSPE) operates on the same inputs available to a physics-based correlation — superheat temperature, droplet volume, and contact angle — and achieves accuracy comparable to standard correlations, confirming that the neural network learns a mapping similar in fidelity to what physics-based models produce from operating condition data alone. The progressive improvement to 5.8% (Case C) and 4.6% (multimodal) demonstrates that the CNN extracts information from image and acoustic data that is fundamentally inaccessible to any correlation based solely on operating condition variables, regardless of how sophisticated the functional form. This distinction is important: the multimodal advantage is not merely a matter of fitting more parameters to the same data, but of accessing additional physical information about the instantaneous state of the boiling process that operating condition variables cannot capture.

The multimodal approach developed in this dissertation is, to our knowledge, the first to combine visual, digital, and acoustic sensing modalities for boiling heat transfer prediction. The demonstrated 20.7% improvement from adding acoustic data to the image-plus-digital baseline provides the first quantitative evidence that acoustic sensing captures complementary information about boiling dynamics that is not available from imaging or digital sensors alone. This finding establishes acoustic emission monitoring as a valuable addition to the experimental toolkit for boiling heat transfer research, particularly for challenging regimes like PDVR, where visual observation alone is insufficient.

5.8. Error Analysis and Sources of Uncertainty

A detailed examination of the prediction errors across all modeling approaches provides physical insight into the sources of remaining uncertainty and the conditions under which multimodal analysis provides the greatest benefit. For the digital-only Case A model, the prediction errors are relatively uniformly distributed across the superheat range, with no strong dependence on regime or operating conditions. This uniform error distribution reflects the fundamental limitation that digital sensor data alone cannot distinguish between different morphological states that may exist at similar operating conditions. For example, at moderate superheats (25–35°C), the boiling morphology can vary significantly between runs depending on the precise locations and timing of bubble nucleation events, but the digital sensor readings (temperature, volume, contact angle) are essentially identical across these runs. The digital-only model therefore predicts the average heat flux for a given set of operating conditions but cannot account for the run-to-run variability associated with morphological differences.

For Case B (image + non-thermal), the prediction errors are largest at high superheats corresponding to the PDVR regime. This is physically expected because the PDVR regime produces the most complex and rapidly evolving two-phase morphology, with chaotic liquid ejection and transient dryout patterns that change faster than the camera frame rate in some cases. The image captured at any single instant may not be representative of the time-averaged process, introducing sampling uncertainty that is inherent to the imaging approach. The acoustic data, which integrates information over the entire vaporization event through the quartile feature extraction, is less susceptible to this instantaneous sampling issue.

For Case C (image + digital), the two misclassified images on the testing dataset occurred at the boundary between the vigorous nucleation and PDVR regimes, where the visual morphology exhibits transitional characteristics of both regimes. The thermal sensor data alone cannot resolve this ambiguity because the superheat temperature ranges of these two regimes overlap in the 35–40°C range. The addition of acoustic data in the multimodal fusion model resolves this boundary ambiguity because the acoustic signatures of vigorous nucleation (characterized by frequent, moderate-amplitude events from bubble nucleation and coalescence) and PDVR (characterized by intermittent, high-amplitude bursts from vapor recoil ejection events) are distinguishable even when the visual morphology appears transitional.

For the multimodal fusion model, the residual 4.6% RMSPE can be attributed to several sources: (1) experimental uncertainty in the heat flux calculation (±5% from uncertainties in droplet volume, wetted area measurement, and evaporation time determination), (2) inherent stochasticity of the boiling process (run-to-run variability in nucleation site locations and bubble dynamics under nominally identical conditions), (3) temporal sampling effects (the image represents an instantaneous snapshot while the heat flux is a time-averaged quantity), and (4) the limited dimensionality of the acoustic feature representation (12 quartile features may not capture all relevant acoustic information). The close match between the model error (4.6%) and the estimated experimental uncertainty (±5%) suggests that the multimodal model has captured nearly all deterministic information content in the combined sensor data, and further improvement would require either reducing experimental uncertainty or fundamentally different measurement approaches.

5.9. The Role of the Skip Connection Architecture

The skip connection is a critical architectural element that warrants detailed analysis. In the hybrid parallel-series CNN architecture developed in this dissertation, the skip connections serve to concatenate the digital sensor data (and acoustic features in the multimodal version) to both output paths of the dual-branch convolutional architecture, providing boundary condition context to both the regime classification and the heat flux prediction. This design choice has both practical and theoretical motivations.

The CNN architectures developed in this dissertation employ standard hyperparameter choices: 32 convolutional filters per layer, 3×3 kernel size, 64 neurons in fully connected layers, ReLU activation throughout, and Adam optimization at a learning rate of 0.001. These values were selected based on established conventions in the CNN literature and preliminary observations during early model development, rather than through systematic hyperparameter optimization. The 32-neuron versus 64-neuron comparison described in Section 3.5 represents the only documented hyperparameter variation, showing diminishing returns beyond 64 neurons that motivated the adoption of 64 as the standard dense layer width across all architectures. Notably, no regularization techniques (dropout, batch normalization, weight decay) were employed, relying instead on the relatively small network size and the data augmentation strategy described in Section 3.3.1 to mitigate overfitting. The deliberate choice to hold architecture and hyperparameters constant across Cases A, B, C, and the multimodal fusion network is methodologically important: it ensures that the progressive improvement from 15.8% to 5.8% to 4.6% RMSPE is attributable entirely to the incorporation of additional data modalities rather than to differences in network capacity or optimization. Systematic hyperparameter optimization — including exploration of deeper architectures, alternative filter sizes, regularization strategies, and learning rate schedules — could potentially improve the absolute performance of each architecture, but would not alter the relative comparison that constitutes the central finding of this dissertation. Such optimization is recommended as future work in Section 6.5.

From a practical perspective, the skip connection prevents the digital data from being "diluted" by the high-dimensional image features. Without the skip connection, the digital data would need to be concatenated with the image data at the input level (early fusion) or processed through separate fully connected layers and combined only at the output level (late fusion). Early fusion risks the digital data being overwhelmed by the much higher-dimensional image data during convolutional processing. Late fusion prevents the model from learning interactions between digital and image features. The skip connection enables intermediate fusion that preserves the identity of the digital data while allowing the fully connected layers after the concatenation point to learn cross-modal interactions.

From a theoretical perspective, the skip connection architecture is related to the residual learning framework introduced in the ResNet architecture [64], which demonstrated that skip connections facilitate gradient flow during training and enable deeper networks to learn effectively. In our architecture, the skip connection serves a different but related purpose: it provides a direct gradient path from the loss function to the digital data inputs, ensuring that the network receives strong learning signals for all input modalities regardless of the depth of the convolutional processing applied to the image data.

The effectiveness of the skip connection is demonstrated by comparing the results with and without this architectural element. During the early stages of this research, preliminary experiments were conducted comparing the skip connection architecture against an early fusion approach in which all inputs (image, digital, and acoustic data) were concatenated at the input level before convolutional processing. The early fusion approach produced noticeably degraded heat flux prediction performance, particularly for Cases B and C where the low-dimensional digital data provides essential constraining information that risks being overwhelmed by the high-dimensional image features during convolutional processing. Although these preliminary results were not systematically documented with the rigor required for quantitative reporting, they informed the architectural design decisions presented in this dissertation and are consistent with the broader deep learning literature showing that intermediate fusion with skip connections outperforms early fusion for inputs of heterogeneous dimensionality. A systematic ablation study comparing early, intermediate, and late fusion strategies is recommended as future work in Section 6.5. The skip connection thus represents not merely an architectural convenience but a fundamental design principle for multimodal CNN architectures that process input data of vastly different dimensionalities.

5.10. Regime Transition Physics Revealed by Multimodal Analysis

The multimodal analysis provides new physical insights into the transitions between boiling regimes that complement the traditional understanding based on boiling curves. The transition from conduction-dominated evaporation to isolated bubble nucleation (onset of nucleate boiling, ONB) occurs at approximately 15–20°C superheat on the ZnO nanostructured surfaces used in this study. The multimodal model captures this transition through a combination of visual indicators (appearance of discrete bubbles in the image) and acoustic indicators (onset of discrete acoustic spikes above the background noise level). The combination of both indicators provides more reliable ONB detection than either alone, particularly for cases where the first nucleation events are small and occur near the edge of the droplet, where visual resolution is limited.

The transition from isolated bubble to vigorous nucleation (approximately 25–30°C superheat) is characterized by a gradual increase in the number of active nucleation sites and the onset of bubble-bubble interactions. The multimodal model captures this transition through the progressive increase in both visual complexity (more bubbles, coalescence events, interfacial disturbances) and acoustic intensity (higher amplitude, broader frequency content, more frequent acoustic events). The smooth nature of this transition in the multimodal feature space explains why both the Case C and multimodal fusion models achieve accurate classification: the combined feature space provides sufficient discriminative information even though individual features (visual or acoustic) may show overlap between these regimes.

The transition from vigorous nucleation to PDVR (approximately 35–45°C superheat) is the most challenging to identify and the most important from a thermal management perspective, as it corresponds to the approach to critical heat flux conditions. The multimodal model captures this transition through the distinctive combination of high-amplitude acoustic bursts (from vapor recoil ejection events) with visual features showing partial surface dryout (dry patches visible within the liquid film). The acoustic signature of PDVR onset, intermittent high-amplitude bursts rather than the continuous moderate-amplitude signal of vigorous nucleation, provides a particularly valuable early warning indicator because the acoustic change precedes the development of visually detectable dry patches in many cases.

These physical insights into regime transitions demonstrate that the multimodal approach is not merely a "black box" prediction tool but provides interpretable information about the underlying physics of the boiling process. The permutation importance analysis (Section 4.8) further supports this interpretation by identifying which features are most important for different prediction tasks, connecting the machine learning model to the physical understanding of the boiling phenomena. Visualization of the convolutional feature maps (Appendix D: CNN Feature Map Visualization) confirms that the CNN learns physically interpretable spatial features, including droplet boundaries, vapor jet structures, and surface texture gradients, rather than operating as an opaque black box.

5.11. Limitations

Several limitations of the present study should be acknowledged to properly contextualize the findings and guide future research. The experimental scope is restricted to a single fluid (deionized water at atmospheric pressure), a single surface type (ZnO nanopillar arrays on copper substrates), and a single heat transfer configuration (sessile droplet vaporization). While the multimodal framework itself is architecture-agnostic and adaptable to other systems, the specific numerical results — accuracy values, feature importance rankings, and regime-dependent trends — are valid only for this particular experimental system. The generalizability of the trained models to different fluids, surface morphologies, or boiling configurations has not been tested and cannot be assumed. The transfer learning approach demonstrated by Lim et al. [63], in which a model pre-trained on pool boiling data was fine-tuned for flow boiling, suggests that adaptation may be feasible, but this remains to be verified for the multimodal framework developed here.

The dataset of 452 images from 123 unique vaporization events is small by deep learning standards, where training sets of thousands to millions of examples are common. The data augmentation strategy (random rotations and flips) partially mitigates this limitation by exploiting the rotational symmetry of the droplet vaporization process, and the consistent performance across training and testing splits suggests the models have not severely overfit. Nevertheless, the small dataset limits the complexity of architectures that can be reliably trained and may contribute to the sensitivity of results to the particular training-testing partition.

Each observation is treated as independent — the CNN receives a single image, a single set of digital values, and a single set of acoustic features without temporal context. The boiling process is inherently dynamic, and the trajectory of how the system arrived at its current state (e.g., whether superheat is increasing or decreasing, whether nucleation sites have recently activated or deactivated) likely contains predictive information that the current framework does not exploit. Temporal architectures such as CNN-LSTM or temporal transformers, as discussed in Section 7.4, could address this limitation.

The acoustic feature representation uses hand-crafted quartile statistics rather than learned representations. While the quartile-based approach is interpretable and computationally efficient, it discards potentially useful information in the raw acoustic signal — particularly fine-grained spectral structure and transient events that do not align with quartile boundaries. End-to-end learning from raw waveforms or spectrograms, as demonstrated by Sinha et al. [61] and Lim et al. [63], could capture richer acoustic information at the cost of increased data requirements and reduced interpretability.

The reported accuracy values (RMSPE, classification accuracy) are based on single training runs with a fixed 85/15 training-testing split, without cross-validation or repeated trials with different random initializations. Neural network training is stochastic, and the reported values therefore represent point estimates rather than statistically robust measures. Multiple training runs with different random seeds would provide confidence intervals on the reported accuracies, and k-fold cross-validation would assess sensitivity to the particular data partition. Nevertheless, the large magnitude of the progressive improvement — from 15.8% to 10.3% to 5.8% to 4.6% RMSPE — and the consistency between training and testing performance across all architectures suggest that the observed multimodal advantage reflects genuine information content rather than stochastic variation in a single training run.

Finally, the multimodal framework has been validated only in an offline, post-processing context using pre-recorded experimental data. Real-time deployment, where inference must occur within the timescale of the boiling process and inform closed-loop control decisions, introduces additional challenges including latency constraints, sensor synchronization, and computational resource limitations that have not been addressed in this work.

5.12. Summary

This chapter has provided a comprehensive assessment of how multimodal information analysis enhances boiling heat transfer research, synthesizing the findings across all CNN architectures developed in Chapters 3 and 4. The progressive improvement trajectory from 15.8% RMSPE (digital only) to 10.3% (image + non-thermal digital) to 5.8% (image + all digital) to 4.6% (multimodal fusion with acoustics) demonstrates that each sensing modality captures fundamentally different physical information about the boiling process: digital sensors capture boundary conditions and thermodynamic state variables, imaging captures spatial morphology and two-phase structure, and acoustic emissions capture temporal dynamics at frequencies and in locations inaccessible to the other modalities.

The 4.6% RMSPE achieved by the multimodal fusion network approaches the estimated experimental uncertainty of ±5–8%, suggesting that the model captures nearly all deterministic information content in the combined sensor data. The 100% regime classification accuracy on 80 test images, compared to 96.9% for image-plus-digital alone, confirms that acoustic data resolves classification ambiguities at regime boundaries, particularly between vigorous nucleation and PDVR where visual morphology is most transitional. The hybrid parallel-series CNN architecture with skip connections provides an effective framework for processing inputs of heterogeneous dimensionality, enabling intermediate fusion that preserves the identity of low-dimensional digital and acoustic data while learning cross-modal interactions with high-dimensional image features. These findings establish multimodal sensor fusion as a powerful paradigm for boiling heat transfer research that extracts substantially more information from each experiment than any single-modality approach.

Chapter 6. Future Directions of Multimodal Modeling – Potential for Widely Diverse Applications

The multimodal sensor fusion framework developed in this dissertation for droplet vaporization on nanostructured surfaces establishes foundational capabilities with broad applicability across heat transfer engineering, energy systems, and autonomous technologies. The progressive improvement from 15.8% to 4.6% RMSPE demonstrated in Chapters 3 through 5 validates the core principle that complementary sensing modalities (visual, acoustic, and digital) capture fundamentally different physical information about the boiling process. This chapter explores the most promising directions for extending this work, organized into three themes: extension to other heat transfer configurations, advancement of the machine learning architecture and sensing hardware, and integration with emerging computational and autonomous technology frameworks.

6.1. Extension to Other Boiling Configurations

The most immediate and impactful extensions of this work involve applying the multimodal CNN framework to boiling configurations beyond deposited droplet vaporization on nanostructured surfaces. Two configurations are particularly promising.

6.1.1. Flow Boiling in Microchannels

Flow boiling in microchannels represents a high-priority application for multimodal machine learning approaches. Microchannels are increasingly used in electronics cooling, fuel cells, and compact heat exchangers due to their high surface-area-to-volume ratios and the resulting high heat transfer coefficients. However, flow boiling in microchannels is characterized by complex and rapidly evolving two-phase flow patterns, including bubbly flow, slug flow, annular flow, and churn flow, that are difficult to predict from operating conditions alone. The spatial confinement of the microchannel means that individual bubble dynamics can significantly affect the overall flow pattern and heat transfer performance, creating a highly stochastic process.

Yang et al. [15] demonstrated that CNN models can identify flow patterns from high-speed camera visualization of flow boiling in microchannels. Extending this to a multimodal framework that incorporates acoustic data could enable detection of flow instabilities and dryout events that precede the visual onset of these phenomena, as acoustic emissions from bubble nucleation and collapse propagate through the channel walls faster than flow pattern changes become visually apparent. The quartile-based acoustic feature extraction developed in Chapter 4 could be directly adapted for flow boiling applications by segmenting the continuous acoustic signal into fixed-duration windows rather than quartiles of the total evaporation time. A key advantage of acoustic sensing in microchannel geometries is that the channel walls themselves serve as acoustic waveguides, potentially amplifying nucleation events that would be difficult to detect in open pool configurations.

6.1.2. Pool Boiling on Extended and Enhanced Surfaces

Pool boiling on extended and enhanced surfaces, including metal foams, porous coatings, and structured surfaces with fins and pins, represents another application where multimodal analysis could provide significant advantages. These surfaces create complex three-dimensional two-phase morphologies that are difficult to characterize visually because bubbles form within the porous structure before emerging from the surface. Acoustic emissions, which propagate through the solid structure and the liquid, could capture nucleation events occurring within the porous medium that are invisible to external cameras.

The work of Calati et al. [11] on water pool boiling in metal foams, which achieved generalized neural network predictions from experimental results, could be significantly enhanced by incorporating acoustic sensing to detect the internal bubble dynamics that govern the heat transfer performance. The demonstration by Dunlap et al. [62] that acoustic emissions during pool boiling can be used for nonintrusive heat flux quantification with 4.5% MAPE provides strong evidence that acoustic sensing is effective for pool boiling configurations. Their finding that acoustic features below 512 Hz are most significant for heat flux prediction suggests that even relatively low-bandwidth acoustic sensors could provide meaningful information when combined with visual and digital data.

6.2. Nuclear Thermal-Hydraulics and Safety

Nuclear reactor thermal-hydraulics represents one of the most critical applications for improved boiling heat transfer monitoring and prediction. In nuclear reactors, the coolant temperature and boiling state directly affect reactor safety, and the approach to critical heat flux (CHF) must be detected and prevented to avoid fuel rod damage. Current monitoring approaches rely primarily on thermocouple measurements at limited locations, which may not capture localized boiling phenomena that can precede system-wide thermal excursions.

Sinha et al. [61] demonstrated that deep learning applied to acoustic spectrograms can predict the approach to boiling crisis with high accuracy, providing advance warning before CHF is reached. Their work showed that peak acoustic frequency shifts systematically from 200–250 Hz during nucleate boiling to 400–500 Hz at CHF onset, providing a clear acoustic precursor to the boiling crisis. A 2024 review in Progress in Nuclear Energy (177:105474) comprehensively documented that ML methods consistently outperform traditional correlations and lookup tables for CHF prediction, with physics-informed hybrid models showing superior extrapolation to conditions outside the training data range. Nuclear CHF monitoring could be transformed by combining traditional thermocouple data with acoustic emission monitoring (since bubble noise changes dramatically near CHF), neutron noise analysis, and gamma densitometry for void fraction measurement.

The quartile-based feature extraction developed in this dissertation is particularly well-suited for nuclear applications because it reduces the acoustic signal to a compact set of summary features that can be processed in real time without the computational overhead of full spectrogram analysis. This efficiency is critical for safety-critical applications where the time between detection and response must be minimized. The multimodal fusion architecture could be trained on experimental data from separate-effects test facilities and then deployed for monitoring in operating reactors, potentially providing continuous, non-intrusive assessment of the boiling state throughout the reactor core. A key advantage of the multimodal approach in nuclear applications is the redundancy it provides: if one sensing modality degrades or fails, the remaining modalities can continue to provide useful predictions, a property demonstrated across all domains reviewed in Chapter 2 and confirmed in this dissertation by the finding that acoustic-only achieves 13.3% RMSPE superheat temperature prediction even without visual or other sensing data.

6.3. Data Center Cooling

The explosive growth of data center computing power, driven by artificial intelligence workloads, has created urgent demand for more efficient cooling solutions. Immersion cooling, where electronic components are submerged in a dielectric fluid that absorbs heat through boiling, has emerged as a promising approach for cooling high-power-density servers. In two-phase immersion cooling systems, the boiling state must be carefully monitored and controlled to maintain optimal heat transfer while preventing dryout conditions that could damage the electronics.

Evans and Gao [65] demonstrated that Google DeepMind's machine learning system achieved a 40% reduction in cooling energy consumption and a 15% reduction in overall Power Usage Effectiveness (PUE) by training deep neural networks on thousands of sensor data points within Google's data centers. More recently, Zhou et al. [66] reported a 20% reduction in supply fan energy consumption and 4% reduction in water usage by applying simulator-based offline reinforcement learning to optimize airflow setpoints in their data center cooling systems. However, both systems primarily rely on point sensor data (temperature, flow rate, pressure) without incorporating imaging or acoustic sensing. The multimodal monitoring framework developed in this dissertation could extend these approaches by adding acoustic sensors mounted on server racks or submerged in the cooling fluid to detect the onset of nucleate boiling, the transition to vigorous boiling, and the approach to critical heat flux conditions, providing early warning of thermal management challenges that point sensors alone cannot detect.

The CNN-as-temperature-sensor concept (Chapter 3, Case B) could enable non-contact temperature monitoring through cameras observing the boiling fluid, complementing traditional temperature sensors that may not capture hot spots on individual electronic components. The demonstrated ability to classify boiling regimes with 100% accuracy using multimodal fusion could enable automated control systems that adjust cooling fluid flow rates, temperatures, or immersion depths in response to detected regime transitions, closing the loop between sensing and control without human intervention.

6.4. Digital Twins and Autonomous Thermal Management

Digital twin technology, creating virtual replicas of physical systems that are continuously updated with real-time sensor data, represents a natural extension of the multimodal monitoring framework developed in this dissertation. Huang et al. [67] developed a deep multimodal information fusion digital twin for aero-engine fault diagnosis, combining sensor measurements with physics-based component-level model simulations via dual Deep Boltzmann Machines mapped into a joint representation space. Chen et al. [68] created a multisensor fusion digital twin for additive manufacturing that combined melt pool images, infrared thermal imaging, and acoustic emission into a location-specific quality prediction system that achieved 96% overall accuracy and outperformed single-sensor approaches, enabling real-time closed-loop feedback control. This architecture directly parallels the challenge of combining visual, acoustic, and thermal data in boiling heat transfer applications.

A digital twin for a boiling heat transfer system could integrate the multimodal CNN model developed in this dissertation with a physics-based simulation of the boiling process. The CNN model would process real-time sensor data (images, acoustic signals, temperature measurements) to provide continuous predictions of heat flux and boiling regime, while the physics-based model would provide constraints based on conservation equations and constitutive relationships. The discrepancy between the data-driven and physics-based predictions could be used to detect anomalies, identify model uncertainties, and trigger alerts when the system operates outside its validated range. The digital twin approach is particularly powerful because it combines the rapid inference capability of the trained CNN model with the extrapolation reliability of physics-based models, addressing the well-known limitation of purely data-driven approaches when encountering conditions outside the training data distribution.

The broader field of smart manufacturing is evolving toward agentic systems where autonomous intelligent agents achieve closed loops of perception, decision-making, and action through multimodal fusion across diverse data modalities. The multimodal boiling analysis framework developed in this dissertation aligns with this vision by demonstrating that the fusion of visual, acoustic, and thermal sensing modalities provides a more complete perception of the boiling process than any single modality alone. Recent advances in end-to-end controllers trained directly from sensor data, without requiring data from existing controllers, suggest a path toward autonomous boiling control systems where the multimodal CNN model developed in this dissertation could serve as the perception module, feeding into a control system that automatically adjusts operating conditions to maintain optimal heat transfer performance. The transfer learning approach demonstrated by Lim et al. [63] — in which a transformer model pre-trained on pool boiling acoustic data was successfully fine-tuned for flow boiling prediction, achieving accuracy comparable to a model trained from scratch on flow boiling data alone — suggests that multimodal boiling models could similarly adapt to new fluid/surface combinations without requiring complete retraining from scratch.

6.5. Advanced Architectures

The multimodal CNN architecture developed in this dissertation, while effective, represents a relatively simple fusion strategy (intermediate concatenation with skip connections) compared to the state-of-the-art architectures reviewed in Chapter 2. Several architectural advances from other domains offer promising directions for improved boiling heat transfer analysis.

Attention-based cross-modal fusion, exemplified by the MCAT transformer architecture [24] in medical imaging and the TransFuser architecture [29] in autonomous driving, could enable the model to dynamically weight the contribution of each modality based on the current conditions. For example, in the conduction-dominated regime where acoustic emissions are minimal, attention mechanisms could automatically increase the weight of image features, while in the PDVR regime where visual observation is compromised, attention could shift toward acoustic features. This adaptive weighting could improve performance across all regimes simultaneously. Lim et al. [63] demonstrated that transformer architectures outperform CNNs for acoustic boiling analysis, achieving less than ±20% error margins for heat flux prediction with spectrogram inputs, suggesting that replacing the fully-connected acoustic branch in our architecture with a transformer encoder could yield significant improvements.

Temporal modeling through recurrent architectures or temporal transformers could capture the evolution of the boiling process over time, rather than treating each observation as independent. The boiling process exhibits temporal correlations, the acoustic and visual characteristics at any moment are influenced by the history of bubble nucleation, growth, and departure, that could be exploited by architectures that model temporal sequences. The hybrid CNN-LSTM architectures that have achieved high accuracy in structural health monitoring [41] and process control applications provide architectural templates for incorporating temporal modeling into the boiling analysis framework. For the droplet vaporization experiments in this dissertation, a temporal model could process the sequence of quartile features (Q1 through Q4) as an ordered time series rather than as independent inputs, potentially capturing the dynamics of how the boiling process evolves during each vaporization event.

Foundation models for industrial sensor data represent an emerging paradigm that could transform the approach to multimodal boiling analysis. Fan et al. [48] developed FISHER, a foundation model pre-trained on 17,000 hours of industrial sound, vibration, voltage, and sensor data that outperformed all baseline models (including 1.2 billion-parameter models) despite having only 5.5–22 million parameters. A critical finding from their work is that performance degrades beyond approximately 100 million parameters due to industrial data quality limitations, suggesting that the heat transfer community should prioritize creating well-curated, standardized multimodal datasets rather than scaling model size. The sub-band modeling approach in FISHER, which supports arbitrary sampling rates, could be adapted for thermal-acoustic signal fusion in boiling applications, enabling a single model to process acoustic data from diverse experimental configurations without retraining.

6.6. Expanded Acoustic Sensing Capabilities

The acoustic sensing approach developed in this dissertation used a single RØDE Wireless ME microphone positioned approximately 3 cm from the heated surface, capturing a spatially-integrated acoustic signal from the entire boiling process. Several extensions of the acoustic sensing capability could significantly enhance the information content of the acoustic data.

Multiple microphone arrays could enable spatial localization of acoustic sources on the heated surface through triangulation techniques analogous to those used in seismic monitoring and structural health monitoring. By deploying an array of three or more microphones at known positions around the boiling surface, the time-of-arrival differences of acoustic events at each microphone could be used to triangulate the location of individual nucleation, coalescence, and vapor recoil events. This spatial information would complement the temporal statistical information captured by the quartile features, enabling the multimodal model to learn not only how intense the acoustic events are and when they occur, but also where they occur on the surface. Cheng et al. [42] demonstrated this principle for structural health monitoring, showing that multiple acoustic emission sensors fused through CNNs improve both detection accuracy and spatial localization of damage.

Higher sampling rate acoustic sensors (approaching MHz frequencies) could capture the high-frequency acoustic emissions associated with nanoscale phenomena on the ZnO nanostructured surface, including the initial stages of bubble nucleation within the nanoporous structure and the capillary-driven rewetting of nucleation sites after bubble departure. These high-frequency events occur at timescales much shorter than the 48 kHz sampling rate used in this study and could provide additional discriminative information about the micro-scale heat transfer mechanisms that are not directly observable by either imaging or lower-frequency acoustic sensors.

Hydrophone-based underwater acoustic sensing, where the sensor is immersed in the liquid phase rather than positioned in air above the surface, could capture acoustic emissions with higher fidelity by eliminating the acoustic impedance mismatch at the liquid-air interface that attenuates the signal reaching an air-coupled microphone. Sinha et al. [61] used immersed hydrophones for boiling acoustic analysis, achieving 99.92% regime classification accuracy. For the droplet vaporization configuration used in this dissertation, a miniature hydrophone embedded in the liquid pool surrounding the test surface could capture acoustic emissions from bubble events occurring beneath the liquid surface that may be partially attenuated before reaching an air-coupled microphone.

6.7. Edge Deployment and Real-Time Processing

The deployment of multimodal machine learning models on edge computing devices, rather than requiring cloud-based inference, is critical for real-time industrial applications where communication latency is unacceptable. Recent advances have demonstrated that powerful multimodal AI models can run on resource-constrained devices while matching or exceeding the performance of much larger cloud-based models [49].

For boiling heat transfer monitoring, edge deployment would enable the multimodal CNN model to process camera, acoustic, and temperature data simultaneously at the point of operation, providing real-time predictions of heat flux and boiling regime without network connectivity. The quartile-based acoustic feature extraction developed in Chapter 4 is particularly well-suited for edge deployment because it reduces the acoustic signal to just 12 summary features that can be computed efficiently on embedded processors, unlike spectrogram-based approaches that require Fast Fourier Transform computation on large data windows. The total network size of approximately 1.3 million parameters (as discussed in Section 4.11) is well within the capability of modern edge computing hardware, including GPU-accelerated embedded systems and neural processing units designed for inference workloads.

The combination of edge-deployed multimodal models with the digital twin approach described in Section 6.4 could enable fully autonomous thermal management systems that continuously monitor the boiling state, predict future conditions, and adjust operating parameters in real time to maintain optimal performance while ensuring safety margins are preserved.

6.8. Toward Standardized Multimodal Boiling Datasets

A significant barrier to progress in machine learning for boiling heat transfer is the lack of standardized, publicly available datasets that enable reproducible benchmarking of different approaches. Unlike fields such as computer vision (ImageNet), natural language processing (GLUE), and autonomous driving (nuScenes), the boiling heat transfer community has no widely accepted benchmark dataset for evaluating machine learning models. The comprehensive review by Carey [17] identified this as a key limitation constraining the field's progress.

The multimodal data collection framework developed in this dissertation provides a template for creating such a dataset. A standardized multimodal boiling dataset should include synchronized high-speed video, acoustic recordings, and digital sensor measurements, along with carefully documented metadata including surface fabrication parameters, fluid properties, experimental uncertainty estimates, and regime labels verified by expert review. The dataset should span a sufficiently wide range of operating conditions (superheat, pressure, surface properties, fluid type) to enable evaluation of model generalization, and should include multiple repetitions at each condition to enable assessment of model sensitivity to the inherent stochasticity of the boiling process.

The creation of such a dataset would enable the boiling heat transfer community to systematically compare different machine learning architectures, including the CNN approach developed in this dissertation, transformer-based approaches, graph neural networks, and physics-informed neural networks, on identical data, accelerating the identification of the most effective approaches for different aspects of the boiling prediction problem. The FISHER foundation model [48] demonstrated that pre-trained models on large industrial sensor datasets can achieve superior performance on downstream tasks, suggesting that a sufficiently large and diverse multimodal boiling dataset could serve as the basis for a boiling-specific foundation model that captures general acoustic, visual, and thermal patterns across diverse boiling configurations.

6.9. Summary

The multimodal sensor fusion framework developed in this dissertation establishes a foundation with broad applicability across heat transfer engineering, energy systems, nuclear safety, data center cooling, and advanced manufacturing. The most immediate extensions involve applying the framework to other boiling configurations (flow boiling in microchannels, pool boiling on enhanced surfaces) and other working fluids. Medium-term directions include expanded acoustic sensing through microphone arrays, higher sampling rates, and immersed hydrophones, as well as integration with digital twin technology for closed-loop autonomous thermal management. Longer-term directions involve the adoption of advanced architectures (attention-based fusion, temporal modeling with CNN-LSTM hybrids, foundation models) that can further improve prediction accuracy and adaptability, and the creation of standardized multimodal boiling datasets that would accelerate progress across the entire research community. The cross-domain principles identified in Chapter 2, complementary information, intermediate fusion, degraded-condition robustness, time-frequency transformation, and physics-informed learning, provide a roadmap for these extensions and ensure that the multimodal approach will continue to advance as both machine learning architectures and sensor technologies evolve.

Chapter 7.
Concluding Remarks

7.1. Summary of Contributions

This dissertation has presented a systematic investigation of machine learning strategies for analyzing boiling heat transfer on ZnO nanostructured surfaces, progressing from single-modality digital data analysis through combined image-and-digital CNN architectures to a multimodal sensor fusion framework incorporating acoustic emission data. The research produced three major contributions to the fields of boiling heat transfer and machine learning for thermal-fluid systems.

The first contribution is the development and validation of hybrid parallel-series CNN architectures with skip connections for simultaneously analyzing high-speed video images and digital sensor data from boiling experiments. Three architectures of increasing complexity were systematically evaluated. Case A established that a digital-only neural network achieves 15.8% RMSPE for heat flux prediction on testing data, representing the fundamental limitation of approaches that rely solely on traditional sensor measurements. Case B demonstrated that incorporating image data alongside non-thermal digital inputs (droplet size and contact angle) reduces the prediction error to 10.3% RMSPE, confirming that visual morphology contains substantial quantitative information about the boiling process that is not captured by digital sensors. Case C combined image data with all available digital inputs and achieved 5.8% RMSPE for heat flux prediction and 96.9% accuracy for regime classification, a 57.4% relative improvement over the digital-only baseline. These results, published in the ASME Journal of Heat and Mass Transfer [5] and the ASME Heat Transfer Summer Conference [18], demonstrate that the hybrid CNN architecture effectively integrates heterogeneous data modalities through the skip connection design, where digital features bypass the convolutional image processing to connect directly with extracted image features at the fully connected layer.

The second contribution is the demonstration that a trained CNN can function as a non-contact temperature sensor. The Case B architecture, which excludes direct temperature measurement as an input, achieved 7.0% RMSPE in predicting surface superheat temperature from image data alone. A dedicated image-only CNN further achieved 10.9% RMSPE for superheat prediction. These results establish that the two-phase morphology visible in high-speed video imagery encodes sufficient thermal state information for a CNN to infer the surface temperature with useful accuracy, opening the possibility of CNN-based temperature sensing in environments where direct thermocouple placement is impractical.

The third contribution is the integration of acoustic emission monitoring into the multimodal framework and the development of the quartile-based acoustic feature extraction methodology. This work, submitted to the ASME 2026 Heat Transfer Summer Conference [19], demonstrated three key findings: (1) acoustic signals alone contain sufficient information to predict surface superheat with 13.3% RMSPE, establishing that boiling acoustic emissions carry quantitative thermal information not directly accessible through thermal sensors; (2) the multimodal fusion of image, digital, and acoustic data achieves 4.6% RMSPE for heat flux prediction and 100% regime classification accuracy, representing a 20.7% relative improvement over the image-plus-digital baseline; and (3) the quartile-based feature extraction provides an efficient and effective representation of acoustic signals that dramatically reduces data dimensionality while preserving the essential physical information content.

7.2. Key Findings

The progressive improvement in heat flux prediction accuracy across the four modeling approaches investigated in this dissertation provides compelling evidence for several key findings. First, multimodal sensor fusion consistently outperforms single-modality approaches. The progression from 15.8% (digital only) to 10.3% (image + non-thermal) to 5.8% (image + digital) to 4.6% (image + digital + acoustic) RMSPE demonstrates that each additional sensing modality contributes unique, complementary information about the boiling process. The improvements are additive and non-redundant: adding image data to digital sensors provides a 63% relative improvement, and subsequently adding acoustic data provides an additional 20.7% relative improvement.

Second, visual morphology and acoustic dynamics capture fundamentally different physical information about the boiling process. The image data captures the spatial distribution of bubbles, liquid films, wetting fronts, and dryout regions at the instant of observation. The acoustic data captures the temporal dynamics of nucleation, bubble growth and collapse, coalescence, and vapor recoil events at timescales that may exceed the camera frame rate. Digital sensors capture the boundary conditions (temperature, volume, contact angle) that constrain the physical system. The permutation importance analysis confirmed that image and acoustic features contribute most to prediction accuracy, with each modality providing different information depending on whether the prediction task is heat flux regression or regime classification.

Third, the multimodal approach provides its greatest advantage in the most challenging boiling regimes. The improvement from adding acoustic data is most pronounced in the vigorous nucleation and PDVR regimes, where the two-phase morphology is most complex and visual observation alone is insufficient to fully characterize the process. In the PDVR regime, chaotic liquid ejection by vapor recoil forces creates visual ambiguity that is resolved by the distinct acoustic signatures of these events. This finding is consistent with the cross-domain principle identified in Chapter 2 that multimodal fusion provides disproportionate benefit when individual sensor modalities degrade.

Fourth, the 4.6% RMSPE achieved by the multimodal fusion network approaches the fundamental accuracy limits imposed by experimental uncertainty and the inherent stochasticity of the boiling process. The experimental uncertainty in heat flux measurement is approximately ±5% (Table 4.1), suggesting that the multimodal model has captured nearly all of the deterministic information content available from the combined measurement suite. Further improvements would likely require either reducing experimental uncertainty or employing fundamentally different measurement approaches (such as spatially-resolved temperature mapping or particle tracking velocimetry).

Fifth, the quartile-based acoustic feature extraction provides a practical and effective approach to incorporating acoustic information into machine learning models for boiling analysis. By reducing each acoustic signal from tens of thousands of waveform samples to just 12 summary features, the quartile approach achieves a dramatic dimensionality reduction that is computationally efficient for both training and inference while preserving the essential statistical characteristics (amplitude level, frequency content, temporal variability) that encode physical information about the boiling process.

7.3. Implications for Boiling Heat Transfer Research

Beyond the specific numerical results, this dissertation provides empirical validation of the five cross-domain principles identified in the Chapter 2 literature review. The first principle — that complementary information drives multimodal advantage — is confirmed by the progressive improvement from 15.8% to 4.6% RMSPE, with permutation importance analysis demonstrating that each modality contributes distinct predictive information. The second principle — that intermediate fusion outperforms early and late fusion — is validated by the skip connection architecture, which processes each modality through modality-specific layers before combining them at a shared fully-connected layer, an intermediate fusion strategy that proved effective without requiring the raw input concatenation of early fusion or the independent model training of late fusion. The third principle — that fusion provides its greatest advantage under degraded conditions — is confirmed by the finding that acoustic data contributes most in the PDVR regime, precisely where visual observation is most impaired by chaotic liquid ejection and rapid morphological changes. The fourth principle — that time-frequency transformation bridges heterogeneous data types — is instantiated by the quartile-based STFT feature extraction, which transforms raw acoustic waveforms into compact statistical features compatible with the digital sensor inputs, enabling unified processing alongside image features. The fifth principle — that physics-informed elements improve performance — is reflected in the cascaded architecture design, where the regime classification output (a physics-motivated discrete variable) informs the heat flux prediction, and in the choice of physically meaningful input features and output targets. The convergence of these principles across domains as diverse as medical imaging, autonomous driving, and boiling heat transfer suggests that the multimodal framework developed here is an instantiation of general information-theoretic principles rather than a technique narrowly tuned to a single experimental system.

The results of this dissertation have several important implications for the broader field of boiling heat transfer research. The demonstration that multimodal machine learning can achieve prediction accuracy approaching experimental uncertainty limits suggests that this approach should be adopted as a standard tool in experimental boiling research. The addition of acoustic sensing instrumentation to existing experimental setups is relatively inexpensive and non-intrusive, and the demonstrated 20.7% improvement in prediction accuracy justifies the modest additional complexity.

The CNN-as-temperature-sensor capability has implications for boiling monitoring in industrial systems where direct temperature measurement near the boiling surface is impractical or impossible. In nuclear reactor fuel assemblies, high-temperature chemical reactors, and compact heat exchangers, the ability to infer surface temperature from visual observation alone (7.0% RMSPE) or from acoustic emissions alone (13.3% RMSPE) provides alternative monitoring pathways that do not require physical access to the heated surface.

The multimodal framework provides a template for analyzing other complex thermal-fluid systems where multiple sensing modalities are available. The hybrid parallel-series architecture with skip connections is generally enough to accommodate arbitrary combinations of image data (processed through convolutional layers) and non-image data (processed through dense layers), making it adaptable to other experimental configurations with minimal architectural modification.

7.4. Recommendations for Future Work

Based on the findings of this dissertation, the following directions are recommended for future investigation. Extension of the multimodal framework to flow boiling configurations should be pursued as a priority, as flow boiling is the most common mode of boiling heat transfer in industrial applications and presents additional challenges (flow instabilities, flow pattern transitions, void fraction variation) that could benefit from multimodal analysis. The addition of flow-rate and void-fraction data as additional digital inputs, combined with side-view or cross-sectional imaging and acoustic monitoring, would enable comprehensive characterization of flow boiling phenomena.

Investigation of attention-based cross-modal fusion architectures, following the approaches reviewed in Chapter 2, could improve the model's ability to dynamically weight the contribution of each modality based on the current boiling regime, potentially achieving further improvements in prediction accuracy across the full range of operating conditions.

Development of temporal models (CNN-LSTM or temporal transformer architectures) that capture the time-evolution of the boiling process, rather than treating each observation as independent, could exploit the temporal correlations in the boiling process to achieve improved predictions, particularly for transient phenomena such as regime transitions and the approach to critical heat flux.

Expansion of the acoustic feature extraction beyond quartile statistics to include spectral features (power spectral density, peak frequencies, spectral centroid) and time-frequency features (short-time Fourier transform coefficients, wavelet coefficients) could capture additional information from the acoustic signal that is not fully represented by the quartile summary, particularly at high superheats where the acoustic signal has rich spectral content.

Finally, the creation of a publicly available multimodal dataset for boiling heat transfer, including synchronized high-speed video, acoustic recordings, and digital sensor measurements across multiple boiling configurations, surface types, and working fluids, would accelerate progress in this field by enabling reproducible benchmarking of different machine learning architectures and facilitating the development of foundation models for thermal-fluid sensor data.

Bibliography

[1] Wemp, C.K. and Carey, V.P., 2017. Water wicking and droplet spreading on randomly structured thin nanoporous layers. Langmuir, 33(50), pp.14513-14525.

[2] Kunkle, C.M. and Carey, V.P., 2016, July. Metrics for quantifying surface wetting effects on vaporization processes at nanostructured hydrophilic surfaces. In Heat Transfer Summer Conference (Vol. 50336, p. V002T08A024). American Society of Mechanical Engineers.

[3] Njike, U.T., Cabrera, S., McClure, E.R. and Carey, V.P., 2021, June. Use of Machine Learning Tools to Assess Surface Dryout During Nucleate and Transition Boiling on Surfaces With Different Wetting and Substrate Properties. In Heat Transfer Summer Conference (Vol. 84874, p. V001T11A006). American Society of Mechanical Engineers.

[4] Wemp, C.K. and Carey, V.P., 2019. Heat transport for evaporating droplets on superhydrophilic, thin, nanoporous layers. International Journal of Heat and Mass Transfer, 132, pp.34-51.

[5] Carey, V.P., Wemp, C.K., McClure, E.R. and Cabrera, S., 2018, November. Mechanism interaction during droplet evaporation on nanostructured hydrophilic surfaces. In ASME International Mechanical Engineering Congress and Exposition (Vol. 52118, p. V08AT10A029). American Society of Mechanical Engineers.

[6] Njike, U.T., Silva, A. and Carey, V.P., 2025. A Convolution Neural Network Design for Combined Image and Sensor Data Analysis to Determine Droplet Vaporization Regime and Heat Transfer Performance. ASME Journal of Heat and Mass Transfer, 147(11), p.111602.

[7] McClure, E.R. and Carey, V.P., 2021. Genetic algorithm and deep learning to explore parametric trends in nucleate boiling heat transfer data. Journal of Heat Transfer, 143(12), p.121602.

[8] Silva, A.D. and Carey, V.P., 2023, July. Use of a Genetic Algorithm to Model the Interaction of Conduction and Nucleate Boiling Mechanisms During Evaporation of Water Droplets on Superheated ZnO Nanostructured Surfaces. In Heat Transfer Summer Conference (Vol. 87165, p. V001T02A005). American Society of Mechanical Engineers.

[9] Sajjad, U., Hussain, I. and Wang, C.C., 2021. A high-fidelity approach to correlate the nucleate pool boiling data of roughened surfaces. International Journal of Multiphase Flow, 142, p.103719.

[10] Qiu, Y., Garg, D., Zhou, L., Kharangate, C.R., Kim, S.M. and Mudawar, I., 2020. An artificial neural network model to predict mini/micro-channels saturated flow boiling heat transfer coefficient based on universal consolidated data. International Journal of Heat and Mass Transfer, 149, p.119211.

[11] Calati, M., Righetti, G., Doretti, L., Zilio, C., Longo, G.A., Hooman, K. and Mancin, S., 2021. Water pool boiling in metal foams: From experimental results to a generalized model based on artificial neural network. International Journal of Heat and Mass Transfer, 176, p.121451.

[12] Kuang, Y., Han, F., Sun, L., Zhuan, R. and Wang, W., 2021. Saturated hydrogen nucleate flow boiling heat transfer coefficients study based on artificial neural network. International Journal of Heat and Mass Transfer, 175, p.121406.

[13] Swain, A. and Das, M.K., 2014. Artificial intelligence approach for the prediction of heat transfer coefficient in boiling over tube bundles. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 228(10), pp.1680-1688.

[14] Kishor, N. and Das, M.K., 2010. Soft computing techniques for prediction of boiling heat transfer coefficient of liquids on copper-coated tubes. Applied Artificial Intelligence, 24(3), pp.210-232.

[15] Yang, B., Zhu, X., Wei, B., Liu, M., Li, Y., Lv, Z. and Wang, F., 2023. Computer vision and machine learning methods for heat transfer and fluid flow in complex structural microchannels: A review. Energies, 16(3), p.1500.

[16] Lee, M., 2022. Optimized Experimental Observation and Measurements of Pool Boiling Heat Transfer Using Computer Vision Techniques (Master's thesis, University of California, Irvine).

[17] Carey, V.P., 2025. Machine Learning Enhancement of Heat Transfer Research and Technology Development. Annual Review of Heat Transfer, 28.

[18] Tchouteng Njike, U., Silva, A. and Carey, V.P., 2024, July. A Convolution Neural Network Design for Combined Image and Sensor Data Analysis to Determine Droplet Vaporization Regime and Heat Transfer Performance. In Heat Transfer Summer Conference (Vol. 87905, p. V001T03A005). American Society of Mechanical Engineers.

[19] Tchouteng Njike, U., Silva, A., and Carey, V.P., 2026, "Machine Learning Based Multimodal Sensor Data Analysis for a Boiling Heat Transfer Experiment," paper SHTC2026-183942, Proceedings of the ASME 2026 Heat Transfer Summer Conference, Bellevue, WA [Pending Review].

[20] Zhao, F., Zhang, C., and Geng, B., 2024, "Deep Multimodal Data Fusion," ACM Computing Surveys, 56(9), 216.

[21] Chen, R.J., Lu, M.Y., Williamson, D.F., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B. and Mahmood, F., 2022. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer cell, 40(8), pp.865-878.

[22] Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I. and Mahmood, F., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE transactions on medical imaging, 41(4), pp.757-770.

[23] Vanguri, R.S., Luo, J., Aukerman, A.T., Egger, J.V., Fong, C.J., Horvat, N., Pagano, A., Araujo-Filho, J.D.A.B., Geneslaw, L., Rizvi, H. and Sosa, R., 2022. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer. Nature cancer, 3(10), pp.1151-1164.

[24] Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M. and Mahmood, F., 2021. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4025).

[25] Hemker, K., Simidjievski, N. and Jamnik, M., 2024. HEALNet: multimodal fusion for heterogeneous biomedical data. Advances in Neural Information Processing Systems, 37, pp.64479-64498.

[26] Huang, S.C., Pareek, A., Seyyedi, S., Banerjee, I. and Lungren, M.P., 2020. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ digital medicine, 3(1), p.136.

[27] Tariq, Z., Shah, S.K. and Lee, Y., 2022. Feature-based fusion using CNN for lung and heart sound classification. Sensors, 22(4), p.1521.

[28] Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D.L. and Han, S., 2023, May. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In 2023 IEEE international conference on robotics and automation (ICRA) (pp. 2774-2781). IEEE.

[29] Liang, T., Xie, H., Yu, K., Xia, Z., Lin, Z., Wang, Y., Tang, T., Wang, B. and Tang, Z., 2022. Bevfusion: A simple and robust lidar-camera fusion framework. Advances in neural information processing systems, 35, pp.10421-10434.

[30] Nabati, R. and Qi, H., 2021. Centerfusion: Center-based radar and camera fusion for 3d object detection. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1527-1536).

[31] Chitta, K., Prakash, A., Jaeger, B., Yu, Z., Renz, K. and Geiger, A., 2022. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. IEEE transactions on pattern analysis and machine intelligence, 45(11), pp.12878-12895.

[32] Liu, Z., Chi, C., Cousineau, E., Kuppuswamy, N., Burchfiel, B. and Song, S., 2024. Maniwav: Learning robot manipulation from in-the-wild audio-visual data. arXiv preprint arXiv:2406.19464.

[33] Chen, C., Jain, U., Schissler, C., Gari, S.V.A., Al-Halah, Z., Ithapu, V.K., Robinson, P. and Grauman, K., 2020, August. Soundspaces: Audio-visual navigation in 3d environments. In European conference on computer vision (pp. 17-36). Cham: Springer International Publishing.

[34] Petrich, J., Snow, Z., Corbin, D. and Reutzel, E.W., 2021. Multi-modal sensor fusion with machine learning for data-driven process monitoring for additive manufacturing. Additive Manufacturing, 48, p.102364.

[35] Shevchik, S.A., Kenel, C., Leinenbach, C. and Wasmer, K., 2018. Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. Additive Manufacturing, 21, pp.598-604.

[36] Xia, M., Shao, H., Williams, D., Lu, S., Shu, L. and de Silva, C.W., 2021. Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning. Reliability Engineering & System Safety, 215, p.107938.

[37] Jiang, Q., Wang, Y., Kong, Y., Liu, Y. and Ma, C., 2025. A monitoring network SIMNet for weld penetration status based on multimodal fusion. Scientific Reports, 15(1), p.23355.

[38] Kullu, O. and Cinar, E., 2022. A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults. Machines, 10(11), p.1105.

[39] Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B. and Chanussot, J., 2022. Deep learning in multimodal remote sensing data fusion: A comprehensive review. International Journal of Applied Earth Observation and Geoinformation, 112, p.102926.

[40] Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. and Beroza, G.C., 2020. Earthquake transformer, an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nature communications, 11(1), p.3952.

[41] Hu, Y., Zhang, Q., Zhu, H., Wang, B., Xiong, H. and Wang, H., 2025. Scalable intermediate-term earthquake forecasting with multimodal fusion neural networks. Scientific Reports, 15(1), p.9748.

[42] Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J. and Gao, J., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote sensing of Environment, 241, p.111716.

[43] Shi, B., Hsu, W.N., Lakhotia, K. and Mohamed, A., 2022. Learning audio-visual speech representation by masked multimodal cluster prediction. arXiv preprint arXiv:2201.02184.

[44] Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K.V., Joulin, A. and Misra, I., 2023. Imagebind: One embedding space to bind them all. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 15180-15190).

[45] Tsai, Y.H.H., Bai, S., Liang, P.P., Kolter, J.Z., Morency, L.P. and Salakhutdinov, R., 2019, July. Multimodal transformer for unaligned multimodal language sequences. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 6558-6569).

[46] Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C. and Zong, Y., 2023. A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. Entropy, 25(10), p.1440.

[47] Liu, F., Fu, Z., Wang, Y. and Zheng, Q., 2025. TACFN: transformer-based adaptive cross-modal fusion network for multimodal emotion recognition. arXiv preprint arXiv:2505.06536.

[48] Azimi, M., Eslamlou, A.D. and Pekcan, G., 2020. Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. Sensors, 20(10), p.2778.

[49] Jia, J. and Li, Y., 2023. Deep learning for structural health monitoring: Data, algorithms, applications, challenges, and trends. Sensors, 23(21), p.8824.

[50] Hassani, S., Dackermann, U., Mousavi, M. and Li, J., 2024. A systematic review of data fusion techniques for optimized structural health monitoring. Information Fusion, 103, p.102136.

[51] Yan, J., Liao, J.B., Gao, J.Y., Zhang, W.W., Huang, C.M. and Yu, H.L., 2023. Fusion of audio and vibration signals for bearing fault diagnosis based on a quadratic convolution neural network. Sensors, 23(22), p.9155.

[52] Singh, S.K., Sharma, D., Review of pool and flow boiling heat transfer enhancement through surface modification, Int. J. Heat and Mass Transfer, 181, 2021, 122020.

[53] Carey, V.P., Liquid-Vapor Phase-Change Phenomena, Third Edition, CRC Press, New York, NY, ISBN 978-1-49871-661-1, 2020.

[54] LeCun,Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., L., Jackel, D., Backpropagation Applied to Handwritten Zip Code Recognition, in Neural Computation, 1, no. 4, pp. 541-551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

[55] Biswas, Avishek. "The History of Convolutional Neural Networks for Image Classification (1989-Today)." Towards Data Science, January 17, 2025. https://towardsdatascience.com/the-history-of-convolutional-neural-networks-for-imageclassification-1989-today-5ea8a5c5fe20/.

[56] Wemp, C. and Carey, V., Tuning superhydrophilic nanostructured surfaces to maximize water droplet evaporation heat transfer performance, ASME Journal of Heat Transfer, 140 (10) (2018) 102401–10. doi:10.1115/1.4040142.

[57] Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of big data, 6(1), pp.1-48.

[58] Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).

[59] Leighton, T.G. and Walton, A.J., 1987. An experimental study of the sound emitted from gas bubbles in a liquid. European Journal of Physics, 8(2), pp.98-104.

[60] Prosperetti, A., 1988. Bubble-related ambient noise in the ocean. The Journal of the Acoustical Society of America, 84(3), pp.1042-1054.

[61] Sinha, K.N.R., Kumar, V., Kumar, N., Thakur, A. and Raj, R., 2021. Deep learning the sound of boiling for advance prediction of boiling crisis. Cell Reports Physical Science, 2(3).

[62] Dunlap, C., Pandey, H., Weems, E. and Hu, H., 2023. Nonintrusive heat flux quantification using acoustic emissions during pool boiling. Applied Thermal Engineering, 228, p.120558.

[63] Lim, D., Liu, Y. and Bang, I.C., 2025. Predicting boiling heat flux, heat transfer coefficient, and regimes Non-intrusively using external acoustics and deep learning. Scientific Reports, 15(1), p.22690.

[64] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[65] Evans, R. and Gao, J., 2016. DeepMind AI reduces Google data centre cooling bill by 40%. Google DeepMind Blog, July 20. Available at: https://deepmind.google/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/

[66] Zhou, C., Gao, D., Rivalin, L., Grier, A., Ramirez, G.A. and Fabian, J., 2024. Simulator-Based Reinforcement Learning for Data Center Cooling Optimization. In Deployable RL: From Research to Practice@ Reinforcement Learning Conference 2024.

[67] Huang, Y., Tao, J., Sun, G., Wu, T., Yu, L. and Zhao, X., 2023. A novel digital twin approach based on deep multimodal information fusion for aero-engine fault diagnosis. Energy, 270, p.126894.

[68] Chen, L., Bi, G., Yao, X., Tan, C., Su, J., Ng, N.P.H., Chew, Y., Liu, K. and Moon, S.K., 2023. Multisensor fusion-based digital twin for localized quality prediction in robotic laser-directed energy deposition. Robotics and Computer-Integrated Manufacturing, 84, p.102581.

Appendices

Appendix A: Key Equations

A.1 Mean Heat Flux During Droplet Vaporization

The mean heat flux during droplet vaporization is calculated as:

$$\overline{q"} = \frac{\rho V_0 h_{lv} f}{nA}$$

where $\overline{q"}$ is the mean heat flux [W/cm^2], ρ is the density of liquid water [kg/m^3], V_0 is the initial volume of the droplet [m^3], h_{lv} is the latent heat of vaporization of water [J/kg], f is the frame rate of the camera [fps], A is the average wetted area of the droplet footprint during vaporization [cm^2], and n is the number of frames between the droplet making contact with the surface and its complete evaporation. The average wetted area is approximated as one-half the maximum wetted area, A = A_max / 2, based on the assumption that the wetted area decreases approximately linearly from its maximum value to zero as the vaporization process completes.

A.2 Root Mean Square Percentage Error (RMSPE)

The prediction accuracy of each model is evaluated using the root mean square percentage error:

$$RMSPE = \sqrt{\left[\left(\frac{1}{n}\right) \cdot \sum_{i=1}^{n} \left(\frac{(y_i - \hat{y}_i)}{y_i}\right)^2\right]} \times 100\%$$

where n is the number of test samples, y_i is the measured (true) value of the i-th sample, and $\hat{y}_i$ is the predicted value of the i-th sample. RMSPE is preferred over absolute error metrics because it normalizes the error relative to the true value, providing a scale-independent measure of prediction accuracy that is comparable across different ranges of heat flux and superheat temperature.

A.3 Short-Time Fourier Transform (STFT)

The acoustic features are extracted from the STFT magnitude spectrogram. The STFT of a discrete signal s(n) is defined as:

$$STFT\{s(n)\}(m,k) = \Sigma_n\, s(n) \cdot w(n - mH) \cdot e^{-\frac{j2\pi kn}{N}}$$

where m is the time frame index, k is the frequency bin index, w(n) is the window function of length N, H is the hop length (number of samples between successive frames), and N is the FFT window size. In this dissertation, the STFT is computed using n_fft = 128 (window length), hop_length = 32, and the default Hann window provided by the librosa library. This yields a frequency resolution of Δf = f_s / N = 48,000 / 128 = 375 Hz per bin and 65 frequency bins (from 0 to the Nyquist frequency of 24 kHz)

A.4 Quartile-Based Acoustic Feature Definitions

For each temporal quartile Q_j (j = 1, 2, 3, 4) of the trimmed acoustic signal, three features are computed from the STFT magnitude spectrogram:

Mean amplitude:

$$a_j = \left(\frac{1}{(K \cdot M_j)}\right) \cdot \Sigma_k \Sigma_m |STFT(m,k)|$$

where K is the number of frequency bins (65), M_j is the number of time frames in quartile j, and |STFT(m, k)| is the magnitude of the STFT at time frame m and frequency bin k.
Mean peak frequency:

$$f_j = \left(\frac{1}{M_j}\right) \cdot \Sigma_m f(argmax_k |STFT(m,k)|)$$

where f(·) maps the frequency bin index to its corresponding frequency in Hz, and $argmax_k$ identifies the frequency bin with maximum magnitude in each time frame.
Peak frequency standard deviation:

$$\sigma_j = \sqrt{\left[\left(\frac{1}{M_j}\right) \cdot \Sigma_m (f(argmax_k |STFT(m,k)|) - f_j)^2\right]}$$

which captures the variability of the dominant acoustic frequency across time frames within the quartile.

A.5 Minnaert Frequency

The natural oscillation frequency of a spherical gas bubble in a liquid is given by the Minnaert relationship:

$$f_0 = \frac{1}{2\pi R}\sqrt{\frac{3\gamma P_0}{\rho}}$$

where f_0 is the natural frequency [Hz], R is the equilibrium bubble radius [m], γ is the ratio of specific heats of the gas ($\gamma \approx 1.4$ for air), P_0 is the ambient pressure [Pa], and ρ is the density of the surrounding liquid [kg/m^3]. For air bubbles in water at atmospheric pressure, this simplifies to approximately $f_0 \approx 3.26 / R$, where f_0 is in Hz and R is in meters. This inverse relationship between frequency and bubble radius means that smaller bubbles produce higher-frequency acoustic emissions, providing the physical basis for using acoustic frequency as an indirect measure of bubble size in the quartile-based feature extraction.

A.6 Heat Flux Uncertainty Propagation

The relative uncertainty in the calculated heat flux is estimated by standard error propagation. Since

$$\overline{q"} = \frac{\rho V_0 h_{lv}}{At}$$

and the material properties ρ and h_{lv} have negligible uncertainty:

$$\left(\frac{\delta\overline{q''}}{\overline{q''}}\right)^2 = \left(\frac{\delta V_0}{V_0}\right)^2 + \left(\frac{\delta A_{avg}}{A_{avg}}\right)^2 + \left(\frac{\delta t}{t}\right)^2$$

The individual contributions with quantified uncertainties are:

Droplet volume: δV_0= ±0.3 µL (pipette calibration) on droplets of 5–8 µL, contributing ±4–6% relative uncertainty.
Evaporation time: δt ≈ ±2 frames at 1200 fps, contributing <±1% for typical events spanning hundreds of frames.
Average wetted area: A_avg is approximated as A_max/2, assuming a linear decrease in wetted area over the vaporization duration. The uncertainty in this approximation has not been independently quantified and represents the dominant uncharacterized source of error in the heat flux calculation. Quantification would require frame-by-frame wetted area tracking across multiple test cases.

The quantifiable instrument uncertainty from droplet volume and evaporation time alone yields a lower bound of approximately ±4–6% on the total heat flux uncertainty.

APPENDIX B: PYTHON CODE

This appendix provides the Python code for the core components of the multimodal CNN pipeline developed in this dissertation. The scripts are organized in the order they would be executed in a complete reproduction workflow: (B.1) acoustic signal denoising converts raw microphone recordings into clean waveforms; (B.2) quartile-based feature extraction transforms each denoised waveform into 12 physically interpretable acoustic features; (B.3) image data augmentation and normalization expands the training set and scales all inputs to comparable numerical ranges; (B.4) CNN architecture definitions construct the cascaded dual-branch network for Case C and the multimodal fusion variant with acoustic inputs; and (B.5) model compilation and training configuration specifies the optimizer, loss functions, and loss weighting used to train the multi-output network. Together, these scripts reproduce the full pipeline from raw experimental data to trained prediction models.

B.1 Acoustic Signal Denoising

```
import librosa
import soundfile as sf
import noisereduce as nr

def denoise_audio(input_file, output_file):
    """
    Applies non-stationary noise reduction to a raw acoustic
    recording from the droplet vaporization experiment.

    Parameters:
        input_file (str): Path to the raw WAV file (48 kHz)
        output_file (str): Path for the denoised output WAV file
    """
    # Load audio file at native sampling rate (48 kHz)
```

```
    audio_data, sample_rate = librosa.load(input_file, sr=None)

    # Apply non-stationary noise reduction with maximum suppression
    reduced_noise = nr.reduce_noise(
        y=audio_data,
        sr=sample_rate,
        prop_decrease=1.0,   # Maximum noise reduction strength
        stationary=False     # Adapts noise profile across signal
    )

    # Save the denoised audio
    sf.write(output_file, reduced_noise, sample_rate)
```

B.2 Quartile-Based Acoustic Feature Extraction

```
import numpy as np
import librosa

def split_into_quartiles(signal):
    """
    Splits a 1D signal into four equal-length segments (quartiles).
    Remainder samples are distributed to earlier segments.
    """
    length = len(signal)
    chunk_size = length // 4
    remainder = length % 4

    result = []
    idx = 0
    for i in range(4):
        size = chunk_size + (1 if i < remainder else 0)
        result.append(signal[idx : idx + size])
        idx += size
    return result

def extract_quartile_features(y, sr, n_fft=128, hop_length=32):
    """
    Extracts three features from one quartile segment using STFT.

    Parameters:
        y (np.ndarray): Audio signal segment (one quartile)
        sr (int): Sampling rate (48000 Hz)
        n_fft (int): FFT window size (128 samples)
        hop_length (int): Hop length between frames (32 samples)

    Returns:
        mean_amplitude (float): Mean of STFT magnitude spectrogram
        mean_frequency (float): Mean of per-frame peak frequencies (Hz)
        freq_std (float): Std dev of per-frame peak frequencies (Hz)
    """
    # Compute STFT magnitude spectrogram
    stft_result = librosa.stft(y, n_fft=n_fft, hop_length=hop_length)
    magnitude_spectrogram = np.abs(stft_result)

    # Mean amplitude: average across all frequency bins and time frames
    mean_amplitude = np.mean(magnitude_spectrogram)

    # Frequency array corresponding to each row of the spectrogram
```

```
    frequencies = librosa.fft_frequencies(sr=sr, n_fft=n_fft)

    # Per-frame peak frequency: frequency bin with maximum magnitude
    max_magnitude_indices = np.argmax(magnitude_spectrogram, axis=0)
    peak_freqs = frequencies[max_magnitude_indices]

    # Mean and standard deviation of per-frame peak frequencies
    mean_frequency = np.mean(peak_freqs)
    freq_std = np.std(peak_freqs)

    return mean_amplitude, mean_frequency, freq_std

# ---- Main processing loop ----
# For each experimental test case:
for file in audio_files:
    # Load denoised audio
    x, sr = librosa.load('Clean_Audio/Denoised/' + file, sr=None)

    # Trim silence (12 dB below peak)
    trimmed_x, _ = librosa.effects.trim(x, top_db=12)

    # Split into four temporal quartiles
    quartiles = split_into_quartiles(trimmed_x)

    # Extract 3 features per quartile = 12 features total
    features = []
    for q in quartiles:
        amp, freq, std = extract_quartile_features(q, sr)
        features.extend([amp, freq, std])

    # features = [a1, f1, s1, a2, f2, s2, a3, f3, s3, a4, f4, s4]
```

B.3 Image Data Augmentation and Normalization

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
import numpy as np

def datanormalization(X_data, y_data):
    """Normalize image pixels to [0,1] and heat flux by median."""
    return X_data / 255, y_data / np.median(y_data)

# Configure augmentation: rotation, shifts, shear, zoom, horizontal flip
datagen = ImageDataGenerator(
    rotation_range=5,           # Random rotation up to 5 degrees
    width_shift_range=0.1,      # Random horizontal shift up to 10%
    height_shift_range=0.1,     # Random vertical shift up to 10%
    shear_range=0.1,            # Random shear up to 10%
    zoom_range=0.1,             # Random zoom up to 10%
    horizontal_flip=True,       # Random horizontal flip
    fill_mode="nearest"         # Fill mode for newly created pixels
)

# Stack all non-image inputs and targets as labels so augmentation
# keeps the correspondence between images and their scalar features.
# Columns 0-2: digital inputs (superheat, droplet size, contact angle)
# Columns 3-6: acoustic amplitudes (a1, a2, a3, a4)
# Columns 7-10: acoustic peak frequencies (f1, f2, f3, f4)
```

```
# Columns 11-14: acoustic frequency std devs (sigma1, sigma2, sigma3, sigma4)
# Columns 15-18: regime one-hot encoding (4 classes)
# Column 19: heat flux target
train_labels = np.column_stack((
    X2_train, X3_train, X4_train,
    X5_train, X6_train, X7_train, X8_train,
    X9_train, X10_train, X11_train, X12_train,
    X13_train, X14_train, X15_train, X16_train,
    y1_train, y2_train
))

# Generate 25 augmented copies of the full training set
augmented_images = X_train.copy()
augmented_labels = train_labels.copy()
i = 0
for X_batch, y_batch in datagen.flow(
    X_train, train_labels,
    batch_size=len(X_train),
    save_to_dir="Augmented",
    save_prefix="aug",
    save_format="jpeg"
):
    augmented_images = np.append(augmented_images, X_batch, axis=0)
    augmented_labels = np.append(augmented_labels, y_batch, axis=0)
    i += 1
    if i > 24:
        break

# Combine original + augmented, then shuffle
final_images = np.concatenate((X_train, augmented_images), axis=0)
final_labels = np.concatenate((train_labels, augmented_labels), axis=0)
shuffle_indices = np.random.permutation(len(final_images))
final_images = final_images[shuffle_indices]
final_labels = final_labels[shuffle_indices]

# Split labels back into separate arrays for the multi-input model
# Digital sensor inputs
X2_aug = list(final_labels[:, 0])   # superheat temperature
X3_aug = list(final_labels[:, 1])   # droplet size
X4_aug = list(final_labels[:, 2])   # contact angle

# Acoustic amplitude features
X5_aug = list(final_labels[:, 3])   # a1
X6_aug = list(final_labels[:, 4])   # a2
X7_aug = list(final_labels[:, 5])   # a3
X8_aug = list(final_labels[:, 6])   # a4

# Acoustic peak frequency features
X9_aug = list(final_labels[:, 7])   # f1
X10_aug = list(final_labels[:, 8])  # f2
X11_aug = list(final_labels[:, 9])  # f3
X12_aug = list(final_labels[:, 10]) # f4

# Acoustic frequency std dev features
X13_aug = list(final_labels[:, 11]) # sigma1
X14_aug = list(final_labels[:, 12]) # sigma2
X15_aug = list(final_labels[:, 13]) # sigma3
```

```
X16_aug = list(final_labels[:, 14]) # sigma4

# Targets
y1_aug = list(final_labels[:, [15, 16, 17, 18]])  # regime one-hot
y2_aug = list(final_labels[:, 19])                 # heat flux

# Normalize: images to [0,1], heat flux and scalar features by their median
X_data, y2_data = datanormalization(final_images, y2_aug)

# Digital inputs
X2_data = X2_aug / np.median(X2_aug)
X3_data = X3_aug / np.median(X3_aug)
X4_data = X4_aug / np.median(X4_aug)

# Acoustic amplitudes
X5_data = X5_aug / np.median(X5_aug)
X6_data = X6_aug / np.median(X6_aug)
X7_data = X7_aug / np.median(X7_aug)
X8_data = X8_aug / np.median(X8_aug)

# Acoustic peak frequencies
X9_data = X9_aug / np.median(X9_aug)
X10_data = X10_aug / np.median(X10_aug)
X11_data = X11_aug / np.median(X11_aug)
X12_data = X12_aug / np.median(X12_aug)

# Acoustic frequency std devs
X13_data = X13_aug / np.median(X13_aug)
X14_data = X14_aug / np.median(X14_aug)
X15_data = X15_aug / np.median(X15_aug)
X16_data = X16_aug / np.median(X16_aug)
```

B.4 CNN Architecture

```
import tensorflow as tf

def build_multimodal_model():
    """
    Multimodal fusion CNN for droplet vaporization analysis.
    Cascaded architecture: regime classification informs
    heat flux prediction.

    Two separate convolutional branches process the same image:
    - Branch 1: feeds regime classification
    - Branch 2: feeds heat flux prediction

    Both outputs receive skip connections from digital
    and acoustic inputs.
    """
    # --- Inputs ---
    input_img = tf.keras.Input(shape=(100, 100, 1), name="image")
    super_heat = tf.keras.Input(shape=(1,), name="super_heat")
    droplet_size = tf.keras.Input(shape=(1,), name="initial_droplet_size")
    contact_angle = tf.keras.Input(shape=(1,), name="surface_contact_angle")

    # 12 acoustic quartile features as individual inputs
    amp1 = tf.keras.Input(shape=(1,), name="amp1")
    amp2 = tf.keras.Input(shape=(1,), name="amp2")
```

```
amp3 = tf.keras.Input(shape=(1,), name="amp3")
amp4 = tf.keras.Input(shape=(1,), name="amp4")
freq1 = tf.keras.Input(shape=(1,), name="freq1")
freq2 = tf.keras.Input(shape=(1,), name="freq2")
freq3 = tf.keras.Input(shape=(1,), name="freq3")
freq4 = tf.keras.Input(shape=(1,), name="freq4")
std1 = tf.keras.Input(shape=(1,), name="std1")
std2 = tf.keras.Input(shape=(1,), name="std2")
std3 = tf.keras.Input(shape=(1,), name="std3")
std4 = tf.keras.Input(shape=(1,), name="std4")

# --- Convolutional Branch 1 (Regime Classification) ---
Z1 = tf.keras.layers.Conv2D(32, 3, strides=1, padding="same")(input_img)
A1 = tf.keras.layers.ReLU()(Z1)
P1 = tf.keras.layers.MaxPool2D(2, strides=2, padding="same")(A1)
Z2 = tf.keras.layers.Conv2D(32, 3, strides=1, padding="same")(P1)
A2 = tf.keras.layers.ReLU()(Z2)
P2 = tf.keras.layers.MaxPool2D(2, strides=2, padding="same")(A2)
F1 = tf.keras.layers.Flatten()(P2)
D1 = tf.keras.layers.Dense(64, activation='relu')(F1)

# --- Convolutional Branch 2 (Heat Flux Prediction) ---
Z5 = tf.keras.layers.Conv2D(32, 3, strides=1, padding="same")(input_img)
A5 = tf.keras.layers.ReLU()(Z5)
P5 = tf.keras.layers.MaxPool2D(2, strides=2, padding="same")(A5)
Z6 = tf.keras.layers.Conv2D(32, 3, strides=1, padding="same")(P5)
A6 = tf.keras.layers.ReLU()(Z6)
P6 = tf.keras.layers.MaxPool2D(2, strides=2, padding="same")(A6)
F = tf.keras.layers.Flatten()(P6)
D = tf.keras.layers.Dense(64, activation='relu')(F)

# --- Regime Output ---
# Skip connection: digital + acoustic data concatenated
# with Conv Branch 1 features
D1f = tf.keras.layers.concatenate([
    D1, super_heat, droplet_size, contact_angle,
    amp1, amp2, amp3, amp4,
    freq1, freq2, freq3, freq4,
    std1, std2, std3, std4
])
regime = tf.keras.layers.Dense(
    4, activation='softmax', name="regime"
)(D1f)

# --- Heat Flux Output (cascaded from regime) ---
# Regime prediction feeds forward as input, along with
# Conv Branch 2 features and all skip connection data
x = tf.keras.layers.concatenate([
    regime, super_heat, droplet_size, contact_angle, D,
    amp1, amp2, amp3, amp4,
    freq1, freq2, freq3, freq4,
    std1, std2, std3, std4
])
Ds = tf.keras.layers.Dense(64, activation='relu')(x)
heat_transfer_rate = tf.keras.layers.Dense(
    1, activation='linear', name="heat_transfer_rate"
)(Ds)
```

```
    # --- Model ---
    model = tf.keras.Model(
        inputs=[input_img, super_heat, droplet_size, contact_angle,
                amp1, amp2, amp3, amp4,
                freq1, freq2, freq3, freq4,
                std1, std2, std3, std4],
        outputs=[regime, heat_transfer_rate],
    )
    return model
```

B.5 Model Compilation and Training Configuration

```
model.compile(
    optimizer=tf.keras.optimizers.Adam(1e-3),
    loss={
        "regime": tf.keras.losses.CategoricalCrossentropy(from_logits=False),
        "heat_transfer_rate": tf.keras.losses.MeanAbsoluteError(),
    },
    loss_weights=[0.2, 0.8],  # Regression weighted 4x classification
)
```

Appendix C: Experimental Acoustic Feature Data

Table C.1 presents the complete dataset of quartile-based acoustic features extracted from 123 unique droplet vaporization events conducted on ZnO nanostructured surfaces. Each row represents one vaporization event, with the 12 acoustic features (four quartiles × three features per quartile) extracted using the methodology described in Section 4.3.1. The superheat temperature ΔT and initial droplet volume V_0 are the experimental conditions, and q'' is the mean heat flux calculated using Equation 3.1. The 12 cases without heat flux values (marked as blank) correspond to data points used exclusively for testing. The regime label is inferred from the superheat temperature: conduction ($\Delta T \leq 15°C$), isolated bubble ($\Delta T = 20–25°C$), vigorous nucleation ($\Delta T = 30–40°C$), and PDVR ($\Delta T = 45–55°C$). Note that 452 high-speed video images were captured across these 123 vaporization events, as each event produces multiple video frames but a single set of acoustic features computed over the entire vaporization duration.

ΔT (°C)	V_0 (μL)	a_1	a_2	a_3	a_4	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)	f_4 (Hz)	σ_1 (Hz)	σ_2 (Hz)	σ_3 (Hz)	σ_4 (Hz)	q'' (W/cm²)	Regime
10	5	0.001098409	0.005385576	0.002870957	0.001096465	213.0835843	1632.718373	2949.962349	1245.293675	780.9217363	1210.157295	2137.457463	2327.371244		Conduction
10	5	0.002652305	0.005140114	0.002656361	0.001716744	785.9083489	988.1704757	1436.403918	1762.937267	1720.832896	1316.71137	1696.532407	2832.281883	20.12007332	Conduction
10	5	0.003137231	0.00533516	0.002485061	0.000880934	582.421875	1262.369792	685.4166667	80.20833333	1144.504986	1624.796109	1287.894173	245.234805	18.20573625	Conduction
10	6	0.00412463	0.001313317	0.000434451	0.000297109	2145.130635	1197.387295	1536.270492	62.12858607	2524.533619	1592.653883	2465.283708	163.3742735	16.13954384	Conduction
10	6	0.000711003	0.004818285	0.004698819	0.00197972	71.69436416	1402.023121	1799.329389	1357.214144	188.9049962	1604.311898	1799.637127	1811.878048	16.14610619	Conduction
10	7	0.000648423	0.004126748	0.003096803	0.00040477	77.30945122	2161.303354	1212.581936	57.14176829	171.3016838	2140.088642	2137.004619	143.0914909	17.31849372	Conduction
10	7	0.006271665	0.005486855	0.003283358	0.001841264	1322.954963	3157.077206	3077.136949	2506.295956	1580.026011	1693.338933	2262.336043	2340.866267	18.47218665	Conduction
10	7	0.005730192	0.0055954	0.004534462	0.002440397	1391.478924	2937.863372	3120.812137	2840.379724	1411.214781	1549.323439	1451.080711	2173.20504	15.17144687	Conduction
10	8	0.001193579	0.006123778	0.001835696	0.000802192	46.08379777	1335.332902	333.5589172	49.37549761	129.5293066	1877.601799	1093.08558	93.34869719	15.08449137	Conduction
10	8	0.008811002	0.005881808	0.003815727	0.003115432	1959.049954	3577.340899	2713.497947	2910.911839	1354.984067	1666.576014	1518.52742	1426.36458	14.46302021	Conduction
10	8	0.002418849	0.009449149	0.006744232	0.003709002	315.4474432	2590.696023	2416.193182	2174.573864	1196.52364	1673.966811	1970.980867	2423.839405	12.10288557	Conduction

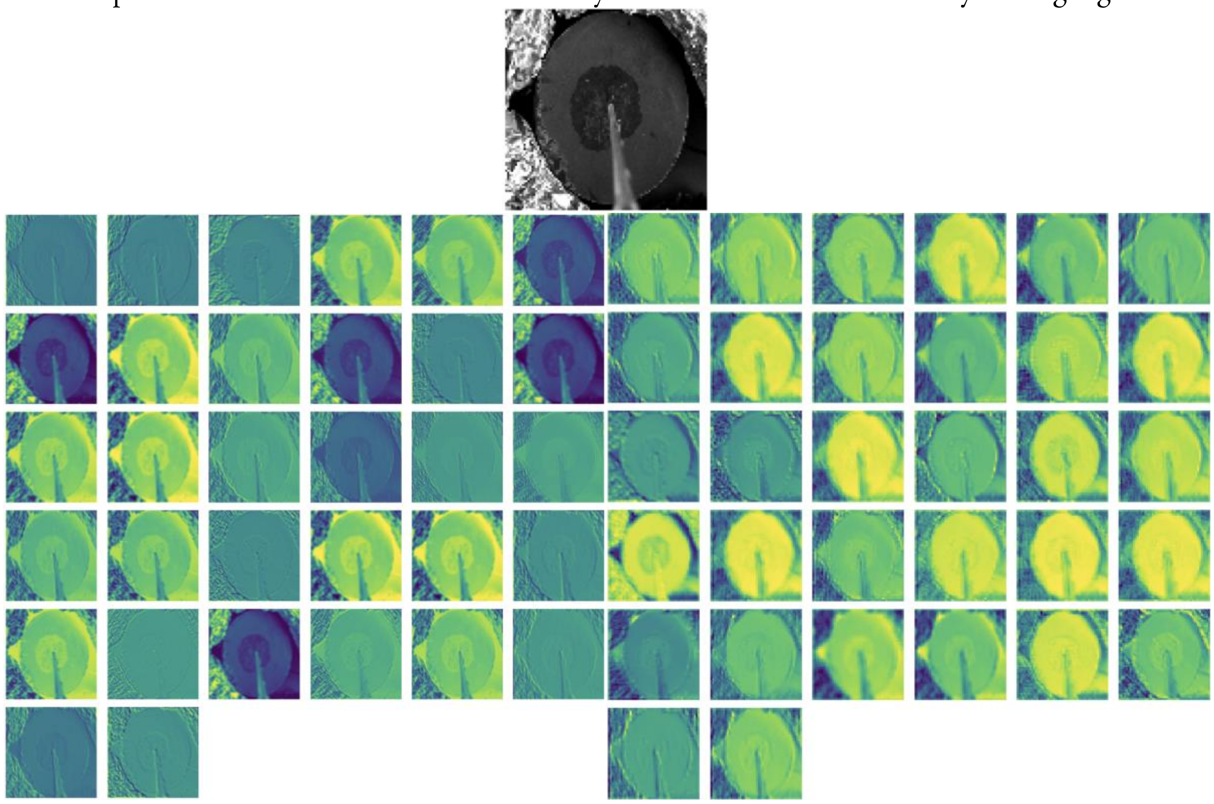
15	5	0.011560486	0.006815942	0.006779394	0.00472085	2091.796875	1729.367898	2771.910511	1991.122159	1563.80912	1693.195591	1678.964585	1306.48812		Conduction
15	5	0.015554686	0.007511591	0.004807104	0.006026826	3252.911932	1272.975852	2015.731534	2346.839489	1961.965648	1578.14612	1540.3065	1714.592114		Conduction
15	5	0.015470761	0.004751123	0.001430576	0.003158238	3652.418364	516.796875	170.3300562	301.948736	2234.038865	1229.54486	756.6366039	759.6213753	18.35872563	Conduction
15	6	0.011132427	0.012285627	0.005795977	0.002881521	4187.109375	3266.015625	1494.140625	1195.3125	1813.162931	1958.250462	1999.695411	1836.462358		Conduction
15	6	0.013558335	0.004628278	0.003991244	0.004154173	2570.339573	1548.685025	1106.934561	1210.976176	1974.824369	1968.388311	1195.298477	1184.212389	16.49728006	Conduction
15	6	0.011771534	0.012521873	0.008552453	0.005169526	3199.892979	1559.829837	1937.39833	3001.669521	1822.623666	1340.712937	1623.719438	1981.012171	16.57993663	Conduction
15	6	0.006968822	0.010639924	0.00584875	0.003110898	1850.46623	2885.912298	2326.512097	1472.593246	2447.559836	2135.319208	2502.701035	2017.585633	16.53769476	Conduction
15	7	0.013875303	0.007056031	0.007499051	0.006862122	2113.495464	1796.748992	1957.900706	2658.077117	1921.283875	2308.75439	2189.930973	1652.969884	14.93184472	Conduction
15	7	0.012527387	0.008069902	0.005810197	0.004249001	3650.919859	2606.212198	3317.50252	2156.09879	2415.982995	2331.150068	2209.81991	2915.751802	13.48319095	Conduction
15	7	0.012522815	0.006120072	0.007508444	0.004711649	3765.763609	2546.938004	2243.157762	2037.550403	2170.289799	2192.270455	1882.670184	2053.96654	14.58887006	Conduction
15	8	0.014339519	0.011863384	0.006741111	0.004931909	2679.393029	1804.813702	1176.706731	1399.326923	1684.339838	1795.865932	1560.975838	1913.146838	16.32153044	Conduction
15	8	0.01474148	0.008272424	0.0037257	0.007178891	2907.248264	2035.286458	712.4565972	1529.123264	2129.584031	2414.771891	1504.641375	1954.853446	13.379288	Conduction
15	8	0.011191516	0.006182034	0.001025594	0.003385258	1944.981971	1104.266827	75.09014423	718.5096154	1477.26672	1386.77625	205.6665467	1046.831491	16.13067384	Conduction
20	5	0.012199683	0.035490669	0.024573831	0.006164709	4223.117898	3889.026989	1727.87642	3226.065341	2599.466682	1858.761891	1391.427597	3144.917674	24.97756688	Isolated Bubble
20	5	0.022520848	0.014955981	0.004946741	0.008875187	2920.043545	1530.622439	1968.346568	1460.021773	1507.411351	1931.184141	2599.743155	2056.477063	21.93034656	Isolated Bubble
20	5	0.01822184	0.036378365	0.020118829	0.007077404	5322.413793	3813.604526	4199.717134	2292.914871	1766.845765	1916.520526	2106.617578	2797.925945	27.92540333	Isolated Bubble
20	5	0.021193374	0.045410909	0.027150011	0.009142144	2756.25	2928.515625	2970.096983	2732.489224	2021.639225	1786.23124	1825.699071	2603.404791	26.33448413	Isolated Bubble
20	6	0.02507809	0.023768431	0.0086308	0.009419365	2242.96875	1968.75	1975.78125	1262.109375	1938.209689	1445.473492	2070.144967	1636.250586	21.52642538	Isolated Bubble
20	6	0.027060742	0.012150384	0.003831846	0.006905395	3984.701588	1389.421107	751.1910861	1745.248463	2107.847884	2227.553834	1713.604844	2065.651461	20.94990224	Isolated Bubble
20	6	0.012024937	0.035468962	0.025929116	0.007853277	1451.207386	2912.855114	3393.110795	2302.09517	1962.770207	1467.47153	2262.617232	2401.187622	23.30394573	Isolated Bubble
20	7	0.016331393	0.034919381	0.025470687	0.005204029	5554.081358	3979.929957	2910.695043	932.6104526	1954.025879	2119.176165	1838.929017	2111.519632	23.13826165	Isolated Bubble
20	7	0.028193273	0.00912273	0.005540385	0.009120121	4389.417614	1671.200284	1776.349432	1380.362216	2228.085096	2427.497821	2292.477227	2120.835843	21.25722402	Isolated Bubble
20	7	0.01331757	0.038139876	0.026873505	0.003266851	3507.954545	3262.606534	2030.646307	1137.997159	2292.362238	1815.644025	1508.800849	1771.523125		Isolated Bubble
20	8	0.019802615	0.045654796	0.03242043	0.00945157	4039.331897	3362.149784	2869.113685	1900.862069	1526.502719	1374.86941	2005.517163	2361.329838	24.15542704	Isolated Bubble
20	8	0.016878473	0.038949367	0.036293834	0.00794998	4865.018858	3403.731142	2797.831358	2465.180496	2084.271774	1737.8291	1649.620848	2581.846386	23.83874992	Isolated Bubble
20	8	0.014684947	0.037454601	0.025687464	0.006452341	3431.345017	3985.388514	3259.079392	1904.23353	2216.731442	2132.938283	1935.077601	2314.28827		Isolated Bubble
25	5	0.020774405	0.058532529	0.026643449	0.002666965	4285.96875	4017.234375	4788.984375	551.25	2620.848109	2046.248799	2486.360557	1551.538637		Isolated Bubble
25	5	0.011413561	0.059528984	0.031634368	0.004266331	2895.703125	4117.96875	3568.359375	3199.21875	2610.708164	1672.633079	2170.652899	2956.404429	42.78001901	Isolated Bubble
25	5	0.015372538	0.044451471	0.015145934	0.009572836	2467.588682	3477.903294	4069.193412	1634.195524	2555.442977	1851.392681	2135.264308	2006.397064	38.30307415	Isolated Bubble
25	5	0.010355754	0.046346359	0.034668922	0.008864848	1734.536638	3843.305496	2756.25	700.9428879	1503.092107	2281.823278	1913.996497	2085.879467	42.44914167	Isolated Bubble
25	6	0.016448146	0.050161779	0.028640643	0.007888107	2387.957974	2595.864763	2358.257004	1116.756466	2160.483742	965.9712494	1895.24772	1640.891897	35.78418765	Isolated Bubble
25	6	0.021698413	0.05257272	0.027327705	0.00838423	2547.443182	3434.872159	1800.958807	876.9886364	1900.256771	1511.886095	1419.963129	2078.624418	26.95692123	Isolated Bubble
25	6	0.021199141	0.053587079	0.021018306	0.004404827	4805.616918	4461.085668	3807.664332	148.5048491	2681.864362	2193.210998	2855.571783	499.1177063	42.70424469	Isolated Bubble
25	7	0.014879629	0.057045035	0.040343907	0.00501375	3154.242996	4680.872845	4033.391703	77.22252155	2769.990854	1524.143112	1915.341831	315.3342329	34.89190812	Isolated Bubble
25	7	0.021209488	0.061528511	0.034913152	0.006101297	2803.771552	3605.697737	3154.242996	1075.175108	2562.668799	1682.344313	2073.170245	1780.29506	39.45001753	Isolated Bubble
25	7	0.009664383	0.049062364	0.048058216	0.005433952	1259.321121	4989.762931	4009.630927	1122.696659	2376.716071	1793.368071	2322.182819	2335.244301	35.80804974	Isolated Bubble
25	8	0.022419488	0.062180646	0.055566829	0.010332793	3486.65625	4554.703125	4148.15625	2742.46875	2742.555314	1922.644903	1682.159495	2672.692151	32.3867156	Isolated Bubble
25	8	0.026429638	0.082869612	0.036319498	0.004710725	4458.234375	2528.859375	2501.296875	468.5625	2590.783368	1432.786133	1970.839212	1168.890497	43.43652157	Isolated Bubble
25	8	0.019347882	0.065330312	0.043902162	0.004623805	2655.266703	4366.042565	4621.470905	273.2489224	2600.570514	1976.737331	2134.919264	727.3766904	32.62087886	Isolated Bubble
30	5	0.003403409	0.069477715	0.045983005	0.007224279	709.734375	4651.171875	3548.671875	2260.125	1494.900409	1655.90192	2285.881111	3240.176741	46.58859213	Vigorous Nucleation
30	5	0.007196837	0.045132808	0.015150802	0.003616854	385.546875	3396.09375	2075.390625	41.015625	348.2224147	1770.507285	2890.13219	73.37098051		Vigorous Nucleation

30	5	0.024360241	0.084872276	0.029109994	0.004185571	3211.03125	5850.140625	3686.484375	55.125	3310.68538	1415.080063	2152.148587	80.3578058	65.48216156	Vigorous Nucleation
30	5	0.022779614	0.102283634	0.060616966	0.01198778	2739.84375	2961.328125	3765.234375	1189.453125	2705.489618	1522.911868	1895.507722	2120.440378	65.05056082	Vigorous Nucleation
30	6	0.012094609	0.09316574	0.051428255	0.008059495	2559.375	3650.390625	3076.171875	237.890625	2941.892701	1774.076324	2016.16726	639.00185	51.44164444	Vigorous Nucleation
30	6	0.004150496	0.081339508	0.044559699	0.005641612	1226.53125	3638.25	2535.75	351.421875	2232.987807	1348.803542	1538.878824	670.695366	41.6054174	Vigorous Nucleation
30	6	0.012899381	0.094697945	0.04720946	0.013153261	2384.15625	4899.234375	3762.28125	2301.46875	2636.364291	1701.217027	2054.908689	2248.118547	64.8948522	Vigorous Nucleation
30	7	0.020340011	0.082553878	0.052869864	0.009351026	2852.71875	4788.984375	6015.515625	3652.03125	2964.60282	2051.625014	1322.676963	3631.823515	44.68539858	Vigorous Nucleation
30	7	0.033894554	0.093871854	0.039161887	0.003410576	4396.21875	5126.625	4478.90625	868.21875	2937.265806	1574.534199	1981.578905	2081.196904	40.08449461	Vigorous Nucleation
30	7	0.043901052	0.099523775	0.049169991	0.014947395	4878.5625	5078.390625	5608.96875	5126.625	2848.888053	1736.464844	1924.496175	2767.973583	58.49525772	Vigorous Nucleation
30	8	0.011467307	0.079813659	0.04581609	0.009045123	1871.161099	3873.006466	3124.542026	1865.220905	2475.6716	1702.712849	2234.981698	2351.161489	49.54678116	Vigorous Nucleation
30	8	0.013363276	0.098726913	0.061807461	0.009553774	2205	2749.359375	3128.34375	806.203125	2436.203832	1077.980968	2145.895957	1516.501174	60.24104675	Vigorous Nucleation
30	8	0.011407182	0.092285	0.078851983	0.007735867	2329.03125	4788.984375	4265.296875	1977.609375	2846.970772	1778.929317	1986.53265	2892.060245	44.78067662	Vigorous Nucleation
35	5	0.006425187	0.082162343	0.039888415	0.002237833	1309.21875	5429.8125	3328.171875	179.15625	2296.760153	764.3936504	2668.158198	509.8131249	88.17511755	Vigorous Nucleation
35	5	0.028851097	0.082953028	0.042614199	0.006835495	5012.109375	4076.953125	2748.046875	1238.671875	2736.649054	1517.999308	1487.097832	1330.778484	57.52119121	Vigorous Nucleation
35	5	0.005137457	0.091603734	0.03426544	0.00692941	1386.328125	5340.234375	4167.1875	1008.984375	2412.58355	1320.524784	1972.472068	1905.174677	59.33209408	Vigorous Nucleation
35	6	0.014329754	0.114255399	0.020185933	0.005129147	2132.8125	5217.1875	3453.515625	926.953125	2703.175521	1360.730063	2222.971199	1739.372024	76.37560529	Vigorous Nucleation
35	6	0.003015452	0.103408419	0.035270467	0.00556253	803.90625	5520.703125	3658.59375	1739.0625	1820.243678	1234.781504	2817.05393	2271.036304	69.34236233	Vigorous Nucleation
35	6	0.023150867	0.109558254	0.054893859	0.010468149	3461.71875	4921.875	4798.828125	1837.5	3180.824524	1138.432986	2780.387474	2612.717848	51.48283699	Vigorous Nucleation
35	7	0.025076155	0.113752797	0.028284099	0.002177385	2921.625	5850.140625	3906.984375	48.234375	2797.491916	913.4692378	2418.9529	91.41447078	79.37315697	Vigorous Nucleation
35	7	0.040340755	0.116833359	0.021027539	0.010436537	4134.375	6431.25	4265.625	3232.03125	3340.777419	936.5224396	2390.475639	3068.210009	85.43705693	Vigorous Nucleation
35	7	0.013793132	0.124058172	0.025626851	0.003572722	1657.03125	5225.390625	3133.59375	713.671875	2501.159582	1485.060183	2079.196886	1559.284386	109.0725484	Vigorous Nucleation
35	8	0.01049017	0.111303896	0.028529217	0.006801746	1770.890625	5974.171875	4795.875	2273.90625	2799.629647	1400.103108	2296.077846	3083.121587	84.83626853	Vigorous Nucleation
35	8	0.002721783	0.095947288	0.063672736	0.006602611	869.53125	5340.234375	5537.109375	1214.0625	2154.534814	1376.072691	1310.088116	1963.033693	99.84044901	Vigorous Nucleation
35	8	0.001671636	0.110161394	0.045859627	0.005933866	668.390625	5581.40625	4878.5625	3252.375	2090.779595	1155.079232	2414.532421	3169.807335	63.45555039	Vigorous Nucleation
40	5	0.00174129	0.099574722	0.019173577	0.009123078	73.828125	5668.359375	2058.984375	1862.109375	85.24937569	847.1091985	2430.091845	2999.136991	134.1098397	Vigorous Nucleation
40	5	0.001736407	0.089128941	0.033932086	0.008190759	186.046875	5023.265625	3314.390625	1323	773.347955	1306.205167	2342.954362	2183.752019	114.9013991	Vigorous Nucleation
40	5	0.001227026	0.093234703	0.022254765	0.006767315	227.390625	4816.546875	2880.28125	358.3125	871.2213704	1409.195989	2494.710789	1477.006714	109.0544247	Vigorous Nucleation
40	6	0.022485793	0.113101497	0.032066036	0.013223735	2288.671875	4995.703125	3232.03125	1460.15625	2634.16126	1404.532154	2843.515991	2226.872717	70.79160025	Vigorous Nucleation
40	6	0.00509792	0.113584258	0.050008334	0.010874454	713.671875	3781.640625	2698.828125	623.4375	2023.862392	1889.676689	2401.933273	1785.079918	89.84026031	Vigorous Nucleation
40	6	0.004891149	0.123617075	0.042852279	0.003666422	1222.265625	5274.609375	5356.640625	689.0625	2505.91707	1287.44643	1975.540049	1687.851526	79.67480644	Vigorous Nucleation
40	7	0.001633908	0.107990406	0.066926137	0.007583347	623.4375	5250	4117.96875	1197.65625	1867.107657	1745.974985	1952.516077	2132.780949	104.8324715	Vigorous Nucleation
40	7	0.001138766	0.115629904	0.042363461	0.010572446	49.21875	4470.703125	4421.484375	1164.84375	77.82167679	1963.753425	1934.650989	2186.713726	75.7652862	Vigorous Nucleation
40	7	0.011923758	0.09026251	0.019630078	0.017190551	1621.672953	4674.932651	2399.838362	3219.585129	2571.287595	1613.919108	2828.779992	3264.470331	70.42970266	Vigorous Nucleation
40	8	0.029830033	0.130760729	0.021212589	0.006345592	2559.375	4519.921875	2715.234375	1000.78125	2681.556268	1215.668806	2190.587665	2243.940816	81.84138934	Vigorous Nucleation
40	8	0.004975013	0.118636556	0.032239594	0.021367099	1353.515625	4618.359375	2460.9375	2501.953125	2305.238679	1954.47953	1547.327704	1937.049469	56.75302337	Vigorous Nucleation
45	5	0.007777587	0.110295288	0.021104001	0.014549542	853.125	4880.859375	2772.65625	533.203125	1750.709085	1243.740921	1371.860763	495.5936825	28.93902332	PDVR
45	5	0.000936751	0.089909427	0.018344633	0.007330479	262.5	4618.359375	2428.125	172.265625	535.9095975	2288.2161	2185.420887	291.1820518	39.56098805	PDVR
45	5	0.011604073	0.096948959	0.024046851	0.010998624	1226.53125	5023.265625	2260.125	1750.21875	2350.56177	1680.578093	2067.94533	2567.014464	36.33234089	PDVR
45	6	0.015505555	0.132662326	0.019036431	0.007818311	2361.05239	5228.768382	2999.448529	1489.590993	3222.12558	1206.497858	2214.809688	2198.896688	54.53271832	PDVR
45	6	0.000982817	0.110690199	0.026356859	0.003871039	65.625	4872.65625	3232.03125	369.140625	124.4065503	1803.717779	2399.915312	687.5960959	49.11555168	PDVR
45	6	0.035256322	0.052856132	0.017031673	0.006637928	2708.015625	2583.984375	1798.453125	1316.109375	3701.272777	1898.364061	2324.541901	2599.729716	39.84116501	PDVR
45	7	0.007733377	0.12870793	0.029335236	0.015508548	1323	4968.140625	3927.65625	2894.0625	2981.626864	1774.70719	3244.452812	3846.425631	42.62950744	PDVR

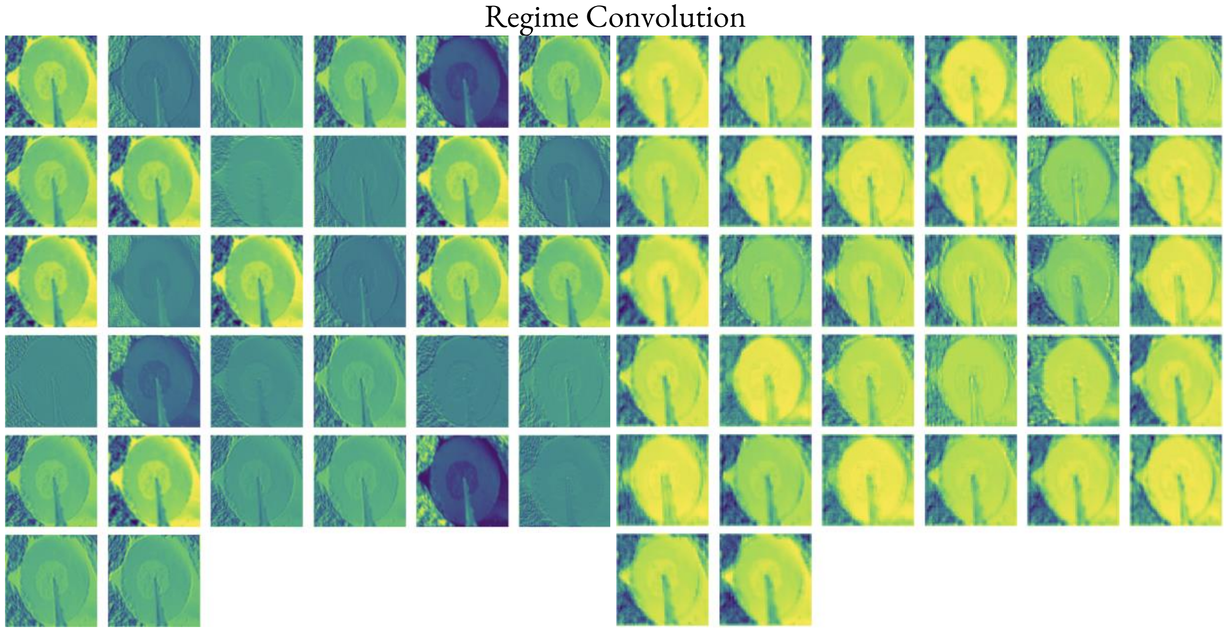
45	7	0.001721513	0.102404349	0.018875174	0.006665897	41.015625	6111.328125	3084.375	98.4375	73.37098051	1207.170047	2415.176093	209.7819913	50.82244284	PDVR
45	7	0.010659027	0.111485943	0.010854974	0.004417333	1706.25	5126.953125	1542.1875	57.421875	2804.485162	1558.939106	1861.043025	81.2067944	48.00854313	PDVR
45	8	0.000943701	0.133977145	0.028434832	0.005833382	57.421875	5996.484375	4708.59375	1008.984375	110.6662066	826.360517	2164.910216	2289.862342	51.39753374	PDVR
45	8	0.002950545	0.102444701	0.010326996	0.006829594	155.859375	5824.21875	1484.765625	311.71875	259.9238827	1201.302775	1402.998193	641.5242355	44.39963497	PDVR
45	8	0.001566878	0.130686179	0.030778363	0.003281347	49.21875	5430.46875	3970.3125	246.09375	77.82167679	1367.439039	2512.433863	704.4187317		PDVR
50	5	0.00831655	0.078575857	0.027110981	0.004085222	1198.96875	5608.96875	4306.640625	530.578125	2224.017937	1896.537351	3005.924095	1557.586115	44.96141704	PDVR
50	5	0.001054854	0.08672522	0.008797091	0.003292574	24.609375	5323.828125	2042.578125	32.8125	60.28041164	1547.110245	2418.322432	67.64470167	44.96141704	PDVR
50	5	0.00720481	0.072882518	0.0063958	0.001472048	1591.40625	5463.28125	1304.296875	49.21875	1587.680884	996.1985239	2317.030705	77.82167679		PDVR
50	5	0.029648967	0.112201557	0.030346546	0.008335993	3182.8125	5742.1875	2953.125	2469.140625	3513.389915	1742.502862	2668.878871	3160.066922	45.76612838	PDVR
50	6	0.002382478	0.076664038	0.011229337	0.004400069	705.46875	6062.109375	1566.796875	131.25	2119.107107	1207.170047	2310.516133	265.3048392	21.73154455	PDVR
50	6	0.015807547	0.109301716	0.008955434	0.015122378	1640.625	5889.84375	2370.703125	3240.234375	3046.353922	730.1242085	2328.185035	3108.800394	25.67240402	PDVR
50	6	0.000761299	0.098533615	0.04331651	0.007652966	0	4749.609375	4897.265625	771.09375	0	2934.334752	2081.881371	1498.592174	29.90734981	PDVR
50	7	0.011382179	0.141180366	0.027426016	0.009962006	2259.719669	5735.431985	3870.909926	1945.588235	3541.690933	1367.65391	2272.430034	2543.309982	44.95705397	PDVR
50	7	0.00120145	0.124638036	0.023410393	0.012812353	16.40625	6554.296875	4429.6875	1722.65625	50.56745862	1105.567023	2902.654656	2611.455535	25.77160177	PDVR
50	7	0.005733112	0.112504095	0.009932879	0.006645017	525	4987.5	1353.515625	713.671875	1480.43113	1854.196603	1839.513057	1043.507147	24.3689517	PDVR
50	8	0.00224759	0.112683192	0.009838455	0.008574584	844.921875	5717.578125	1804.6875	2444.53125	2237.814906	628.1687252	2596.596823	3424.586094	33.36638493	PDVR
50	8	0.046190064	0.061325908	0.014357933	0.009883214	2494.40625	5202.421875	1192.078125	337.640625	3038.179631	1577.336158	2165.849605	643.6006054	31.52905983	PDVR
50	8	0.024457637	0.099798732	0.017922057	0.011402423	1886.71875	5414.0625	4618.359375	1115.625	2999.06968	1499.489965	2493.318219	1964.575653	50.86412749	PDVR
55	5	0.009222141	0.102593161	0.010231065	0.002979008	1104.526654	5623.965993	1712.522978	314.1314338	2307.450849	1427.820824	1748.275782	1130.756932	44.15677002	PDVR
55	5	0.031032568	0.027493645	0.016645849	0.003331071	1452.618243	2067.1875	1401.404139	176.9214527	2558.418627	2733.458609	2446.441427	535.0759488	40.75320559	PDVR
55	5	0.001540742	0.070004493	0.010918341	0.006571952	49.21875	4282.03125	3289.453125	483.984375	77.82167679	2872.874448	2486.670152	893.72661	27.83558882	PDVR
55	6	0.031386379	0.019045474	0.016893588	0.011893896	1534.730114	1696.555398	897.8693182	2589.204545	2563.321634	2082.40968	1623.078513	2959.970141		PDVR
55	6	0.038601585	0.012376634	0.008585236	0.007217349	1997.350084	1811.116976	279.3496622	1182.580236	2781.871663	2320.751482	457.2677039	2286.120446	52.28429694	PDVR
55	6	0.029618125	0.023690909	0.002310528	0.007622368	2183.583193	3547.740709	358.4987331	456.2711149	2958.09617	2811.981179	673.0207898	701.9963078	62.57588988	PDVR
55	7	0.049833979	0.016678069	0.00633281	0.014597326	2028.90625	2074.84375	754.140625	2189.6875	2463.362	1774.681757	2039.014765	3134.9656	59.46693975	PDVR
55	7	0.001397813	0.076944247	0.024236497	0.006090914	41.015625	3026.953125	3108.984375	1911.328125	73.37098051	3133.228668	2320.020105	2967.06022	61.77341464	PDVR
55	7	0.000971233	0.126132891	0.022814453	0.011405085	73.828125	6570.703125	2460.9375	1156.640625	180.8412349	1087.277725	2565.1265	1188.915557	39.6326565	PDVR
55	8	0.05790614	0.008701717	0.010804272	0.012336213	3757.252956	1876.298564	628.5367399	2616.575169	3157.385866	2949.605566	832.6515378	3234.677971	45.11285084	PDVR
55	8	0.065488972	0.014343009	0.016341683	0.002681277	3476.633523	1847.940341	783.0255682	334.0909091	3020.706561	2541.799574	792.3661637	872.1591097	53.50999346	PDVR
55	8	0.001396426	0.10014341	0.016473154	0.002844519	155.859375	5135.15625	4306.640625	369.140625	327.3036595	2197.488426	2463.315258	713.8132942		PDVR

Appendix D: CNN Feature Map Visualization

To provide insight into the learned representations within the convolutional branches, Figure D.1 visualizes the feature maps produced by each of the 32 filters in the two convolutional layers for a representative droplet image from the PDVR regime. In the first convolutional layer (conv2d_24, output shape 100×100×32), the filters extract low-level features including the droplet boundary contour, the vapor jet structure, surface texture gradients, and edge orientations at various angles. The diversity of responses across the 32 filters confirms that the convolutional layer has learned to decompose the boiling image into distinct spatial features relevant to the two-phase morphology. In the second convolutional layer (conv2d_25, output shape 50×50×32), the feature maps become more abstract following max pooling, combining the low-level edge and texture features into higher-level representations of the droplet shape, wetting area, and vapor structures. The progressive abstraction from fine-grained edges to coarse morphological features is consistent with the hierarchical feature learning that characterizes deep convolutional networks. These visualizations demonstrate that the CNN is not operating as an opaque black box but is learning physically interpretable spatial features that correspond to the visual characteristics used by human researchers to identify boiling regimes.



Regime Convolution



HeatFlux Convolution

The Python code used to generate these visualizations is provided below.

```
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt


def visualize_feature_maps(model, input_dict, layer_names=None):
    """
    Visualize feature maps from convolutional layers.

    Args:
        model: Your trained Keras model
        input_dict: Dictionary of inputs (single sample)
        layer_names: List of layer names to visualize (default:
all Conv2D)
    """
    # Find conv layers if not specified
    if layer_names is None:
        layer_names = [layer.name for layer in model.layers
                       if 'conv' in layer.name.lower()]

    # Create a model that outputs feature maps
    outputs   =   [model.get_layer(name).output   for   name   in
layer_names]
    feature_model      =      tf.keras.Model(inputs=model.inputs,
outputs=outputs)

    # Get feature maps
    feature_maps = feature_model.predict(input_dict, verbose=0)

    # Handle single layer case
    if not isinstance(feature_maps, list):
        feature_maps = [feature_maps]

    # Plot each layer's feature maps
    for layer_name, fmaps in zip(layer_names, feature_maps):
        fmaps = fmaps[0]  # Remove batch dimension
        n_filters = min(fmaps.shape[-1], 32)  # Show max 32 filters

        # Grid size
        grid_size = int(np.ceil(np.sqrt(n_filters)))
```

```
        fig,    axes    =    plt.subplots(grid_size,    grid_size,
figsize=(12, 12))
        fig.suptitle(f"Layer:     {layer_name}     —     Shape:
{fmaps.shape}", fontsize=14)

        for i, ax in enumerate(axes.flat):
            if i < n_filters:
                ax.imshow(fmaps[:, :, i], cmap='viridis')
            ax.axis('off')

        plt.tight_layout()
        plt.show()


def visualize_filters(model, layer_name):
    """
    Visualize the learned filter weights themselves.

    Args:
        model: Your trained Keras model
        layer_name: Name of conv layer to visualize
    """
    # Get filter weights
    filters, biases = model.get_layer(layer_name).get_weights()

    # Normalize to 0-1 for visualization
    f_min, f_max = filters.min(), filters.max()
    filters = (filters - f_min) / (f_max - f_min)

    n_filters = min(filters.shape[-1], 32)
    grid_size = int(np.ceil(np.sqrt(n_filters)))

    fig, axes = plt.subplots(grid_size, grid_size, figsize=(8, 8))
    fig.suptitle(f"Filters: {layer_name}", fontsize=14)

    for i, ax in enumerate(axes.flat):
        if i < n_filters:
            # For single-channel input, squeeze the filter
            f = filters[:, :, 0, i] if filters.shape[2] == 1 else
filters[:, :, :, i]
            ax.imshow(f, cmap='gray')
        ax.axis('off')

    plt.tight_layout()
    plt.show()
```

```
#
=================================================================
=============
# for model usage
#
=================================================================
=============

# Prepare a single sample
X_single = {
    "image": (X_val[2:3] / 255).astype(np.float32),
    "super_heat": np.array(X2_val[0:1]),
    "initial_droplet_size": np.array(X3_val[0:1]),
    "surface_contact_angle": np.array(X4_val[0:1]),
    "amp1": np.array(X5_val[0:1]),
    "amp2": np.array(X6_val[0:1]),
    "amp3": np.array(X7_val[0:1]),
    "amp4": np.array(X8_val[0:1]),
    "freq1": np.array(X9_val[0:1]),
    "freq2": np.array(X10_val[0:1]),
    "freq3": np.array(X11_val[0:1]),
    "freq4": np.array(X12_val[0:1]),
    "std1": np.array(X13_val[0:1]),
    "std2": np.array(X14_val[0:1]),
    "std3": np.array(X15_val[0:1]),
    "std4": np.array(X16_val[0:1]),
}

# Show original image
plt.figure(figsize=(4, 4))
plt.imshow(X_val[2].squeeze(), cmap='gray')
plt.title("Original Droplet Image")
plt.axis('off')
plt.show()

# Visualize feature maps from all conv layers
visualize_feature_maps(model2, X_single)

# Visualize learned filters from first conv layer
# visualize_filters(model2, 'conv2d_16')  # Adjust layer name as
needed
```