



Explainable and Low-Power Deep Learning Architectures for Automated Neonatal Central Apnea Detection

Student Name: Mohammad Lakhi

Student ID: 2512430

Subject Code: ACFA7003

University of Wales Trinity Saint David

Date: 12th June 2026

Abstract

In Neonatal Intensive Care Units (NICUs), continuous cardiorespiratory monitoring is crucial. However, conventional threshold-based systems have a high false alarm rate (up to 65%) due to a low event precision floor of 0.35. This causes clinical alarm fatigue, disrupted infant development, and anxious medical teams. To address this issue, this project introduces a multi-parametric machine learning framework that combines synchronized electrocardiogram (ECG), chest impedance, and peripheral oxygen saturation (SpO_2) waveforms to improve central apnea detection. High-frequency vital sign streams are segmented into 30-second moving windows to capture latent cross-channel physiological correlations, and traditional linear classifiers are compared to structurally optimized Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures using a rigorous leave-one-patient-out cross-validation scheme. To transform the system from a non-transparent "black box" into a clinically auditable asset, explicit Explainable AI (XAI) layers via SHapley Additive explanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) are integrated to provide precise quantitative feature weights and visual activation overlays directly onto respiratory time-series signals. To enable point-of-care edge deployment, trained network parameters are converted into 15-bit integers using a 2^{10} scaling factor and synthesized across a 260-cell layout on an Artix-7 FPGA, maintaining an infant apnea detection accuracy above 85% while staying strictly within an ultra-low-power budget under 9W. This architecture establishes a feasible engineering path for embedding intelligent, proactive alerting tools directly inside everyday clinical devices, such as a standard neonatal nasal cannula, effectively reducing clinical alarm noise while maintaining critical infant diagnostic parameters.

TABLE OF CONTENTS

1	Introduction	1
2	Objectives	3
3	Importance of Research	4
4	Literature Review	5
4.1	Threshold-Based Monitoring and Multi-Sensor Baselines	5
4.2	Traditional Machine Learning and Predictive Alarming	5
4.3	Deep Learning Architectures and Validation Deficits	6
4.4	Explainable AI and Bedside Clinician Trust	7
4.5	Hardware Optimization and Edge Processing	7
5	Hypothesis / Research Question	9
6	Research Design and Methodology	10
7	Research Limitations	12
8	Project Timeline	13
9	Expected Findings	14
	REFERENCES	16

Chapter 1

Introduction

For hospitalized newborns in the neonatal intensive care unit (NICU), continuous cardiorespiratory physiological monitoring is an essential component of medical care. Automated monitoring is necessary to monitor the clinical status of preterm newborns because they often have extremely undeveloped cardiorespiratory systems that are vulnerable to unexpected clinical decline. When physiological thresholds are exceeded, traditional bedside patient monitors are set up to sound and display alerts. However, because traditional central apnea monitoring frequently depends on basic chest impedance tracking, typical automatic detection criteria provide a remarkably high proportion of false or clinically irrelevant warnings. Up to 65% of the cardiorespiratory and central apnea alerts generated in active NICUs are incorrect due to this operational technique's inherently low precision of 0.35. Clinical teams are subjected to continuous sensory noise as a result of this systemic impairment, which disturbs newborn sleep and leads to severe clinical alarm fatigue [1, 4].

For the care of infants, there is an urgent medical need to address this ongoing alarm issue. Serious long-term medical disorders in premature newborns, such as neuro developmental impairment, acute late-onset sepsis, and increased patient mortality, are closely correlated with periodic apneas and cardiorespiratory instability. Nurses and caregivers are more likely to overlook real, life-threatening clinical events when bedside monitors continuously set off false alarms. Additionally, bedside practitioners must entirely disable monitor alerting capabilities in response to the disruption of susceptible infants developmental settings caused by the visual and aural disturbance. Therefore, to increase patient safety, lessen workflow burdens, and re-sensitize staff, an automated, low-power, and accurate clinical decision support system is needed [4, 7, 9].

A review of recent scholarly works reveals serious structural and methodological flaws.

A startling 70.7% of all published neonatal intensive care artificial intelligence studies lack any validation process, according to a macro-level systematic review, which raises serious concerns about mathematical overfitting. Furthermore, there are serious issues with external validity and generalizability across diverse patient populations because the great majority of current neonatal monitoring trials are limited to small, single-center retrospective cohorts. East Asian patient cohorts are significantly over-represented in published respiratory datasets due to their strong geographic concentration. Furthermore, there is a distinct lack of cross-institutional collaboration because medical centers rarely retain continuous physiological monitoring data for more than a few days, and the domain lacks uniform interoperability guidelines for sharing complex high-frequency data streams. Current frameworks are also limited by a lack of prospective clinical validation and head-to-head performance comparisons based on standard datasets, benchmarking criteria, or unified outcome definitions. Finally, the growing use of complex deep learning models introduces non-transparent "black-box" architectures that lack physical interpretability and clear explanation tools, clinician trust and regulatory acceptance at the bedside [1, 6, 10].

Chapter 2

Objectives

This study aims to create a low power, non-invasive machine learning system that enhances the accuracy of neonatal cardiorespiratory monitoring.

To achieve this goal, the study outlines three specific and measurable objectives:

- To train and cross-validate traditional and convolutional neural network models using multi-parametric ECG, chest impedance, and oxygen saturation waveforms to raise central apnea detection precision above the conventional 0.35 baseline [4, 9].
- To map quantitative feature contributions for clinical teams, we use explicit explainable artificial intelligence techniques, such as Shapley Additive Explanations and Gradient-weighted Class Activation Mapping[5, 6].
- Validating a compact feedforward network architecture on an FPGA platform involves mapping trained model weights into integer-scaled logic blocks [8].

The study expects to achieve an 80% reduction in acoustic alarms while maintaining an early prediction window for critical de-saturation and bradycardia events [7].

Chapter 3

Importance of Research

The scientific value of completing these objectives is to advance academic knowledge of multiparametric neonatal physiology. This study advances our understanding of continuous central apnea dynamics by identifying latent cross-channel correlations between high-frequency ECG, chest impedance, and oxygen saturation signals [4, 9]. It systematically compares the operational trade-offs of traditional and deep learning architectures, while also introducing quantitative feature validation layers to eliminate non-transparent black-box elements from neonatal predictive analytics [5, 6, 10].

This system effectively combats clinical alarm fatigue by filtering out up to 66% of false alerts in the intensive care unit environment. Shifting monitoring from static rule-based thresholds to proactive prediction windows gives nursing teams more time to protect infants from prolonged desaturation and bradycardia episodes [7].

Furthermore, validating an integer-scaled network layout establishes a clear engineering path toward small-scale, ultra-low-power biomedical chip designs that can fit inside a nasal cannula, allowing for seamless wearable point-of-care diagnosis in the NICU and at home [3, 8].

Chapter 4

Literature Review

4.1 Threshold-Based Monitoring and Multi-Sensor Baselines

Continuous cardiorespiratory monitoring is the foundation for protecting high-risk infants from sudden physiological deterioration in the neonatal intensive care unit(NICU). Traditional baseline approaches for tracking infant health rely heavily on multi-sensor IOT frameworks. These configurations typically include peripheral components such as the MAX30100 sensor for heart rate and blood oxygen saturation, the LM35 sensor for body temperature, load cells for intravenous fluid level monitoring, and cameras to monitor movement or posture . While these multi-sensor systems successfully automate data visualisation and send basic alerts to mobile applications via Python algorithms, they only perform actions by comparing variables to rigid, single-parameter thresholds [2]. As a result, they are completely incapable of recognising latent temporal patterns, resulting in low precision, a high rate of false alarms, and severe clinical alarm fatigue [1, 4].

4.2 Traditional Machine Learning and Predictive Alarming

To overcome traditional threshold limitations, clinical research has shifted to machine learning models that use multi-parametric signal fusion. To identify true central apneas, researchers extracted multi-channel features from electrocardiograms (ECG), chest impedance(CI), and SpO2 waveforms across overlapping 30-second moving windows . Compared to traditional classifiers, Elastic Net Logistic Regression(ENLR)

models fed with all moving window features achieve a top mean Area Under the Receiver Operating Characteristic Curve(AUROC) of 0.88 under leave-one-patient-out cross-validation. Log odds ratio analyses demonstrate that respiration features derived from chest impedance make the most significant mathematical contributions to overall model success [9].

Setting an optimal false positive rate threshold of 0.3 increases baseline clinical precision to 0.42 and correctly detects 95.2% of critical central apneas accompanied by desaturation and bradycardia[9]. At the subcritical level, boosted decision trees can be used to extend this multi-parametric predictive strategy. This framework predicts critical desaturation alerts 18.4 seconds in advance, allowing for early clinical actions that reduce infant disease burdens while cutting total auditory alarms by nearly 80% [7].

4.3 Deep Learning Architectures and Validation Deficits

The literature demonstrates a clear structural transition from traditional statistical models to sophisticated deep learning networks. Macro-level statistical taxonomies show that traditional machine learning algorithms account for 63% of neonatal cardiorespiratory monitoring studies, while deep learning architectures account for 35% and are steadily increasing [3].

When tested on identical clinical datasets, deep neural networks outperform linear baselines, with Convolutional Neural Networks(CNN) and Long Short-Term Memory(LSTM) models maximising event classification. A structurally optimised CNN with a consecutive window detection criterion achieves a central apnea precision of 0.54 and an overall sensitivity of 0.75, demonstrating high technical accuracy when tracking clinical apnea-bradycardia pairings [4].

However, severe methodological deficiencies remain across these advanced architectures. A macro-level systematic review reveals that 70.70% of all published neonatal intensive care artificial intelligence papers do not include any validation procedure, significantly increasing the risk of model overfitting[10]. Furthermore, the current

literature is heavily dominated by retrospective, single-center designs that employ geographically concentrated datasets with an over-representation of East Asian cohorts, severely limiting generalisability and external validation across diverse NICU populations [6].

4.4 Explainable AI and Bedside Clinician Trust

Because complex deep learning models use non-transparent layouts, the lack of feature interpretability is a significant barrier to real-world clinical implementation. Bedside practitioners strongly prefer interpretable and explainable AI models over black-box systems to ensure safety and ethical accountability in high-risk situations [3, 6].

To address this, researchers are incorporating Explainable AI (XAI) layers into neonatal time-series workflows to identify hidden physiological patterns. Methodologies such as SHapley Additive ExPlanations (SHAP) and Integrated Gradients compute and visualise the precise quantitative contributions of individual vital signs to model outputs, directly supporting high-stakes clinical decisions such as caffeine weaning or discharge planning [5]. Furthermore, visualization-based methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) are applied to neural network breath classification tasks, providing clinicians with a clear visual explanation tool to compare algorithmic decisions to real clinical relevance [6].

4.5 Hardware Optimization and Edge Processing

The final dimension of neonatal monitoring entails translating these mathematical models into embedded edge hardware in order to create low-cost on-chip point-of-care devices. Processing high-frequency floating-point vital signs on physical chips increases both hardware resource utilisation and computational overhead. To address this issue, researchers created parameter-conversion methods that multiply signed floating-point parameters by 2^{10} and discard fractional pieces, resulting in compact 15-bit integers [8].

On an Artix-7 FPGA, this integer-scaled logic structure maps 260 digital cells to a compact feedforward neural network configuration. This implementation achieves an

apnea detection accuracy rate of more than 85% while using a total typical power consumption budget of 8.713W. This micro-scale engineering footprint eliminates bulky bedside monitors, allowing fully integrated, low-power machine learning systems to be housed within comfortable clinical tools such as a standard nasal cannula [8].

Chapter 5

Hypothesis / Research Question

A critical evaluation of current neonatal monitoring methodologies reveals that traditional threshold configurations fail to capture latent temporal dependencies [2], while advanced deep architectures introduce non-transparent black-box layers [4, 6]. To address these gaps, this study proposes two research questions and a core hypothesis that are directly related to the stated objectives and clinical alarm fatigue problems:

- **RQ1 (Statistical):** Can a multi-parametric CNN architecture processing fused ECG, chest impedance, and SpO_2 signals increase central apnea precision above the conventional 0.35 floor while reducing false alerts [4, 9].
- **RQ2 (Logical):** Can embedding SHAP and Grad-CAM layers extract visual, decision-level explanations that match human expert reasoning while maintaining baseline model metrics [5, 6].
- **Hypothesis (Analytical):** An asymmetric, integer-scaled network layout compiled across 260 digital cells on an Artix-7 FPGA maintains an overall apnoea detection accuracy above 85% within an 8.713W typical power budget [8].

These formulations can be logically verified through statistical AUROC and precision-recall audits, quantitative feature-importance mapping, and direct physical hardware power simulation measurements [4, 5, 6, 8, 9].

Chapter 6

Research Design and Methodology

This study presents a non-invasive, multi-stage computational methodology for processing multi-parametric physiological data and validating an edge-hardware alerting platform [8, 9]. The steps required to achieve the research objectives are organised as follows:

- **Phase 1: Multiparametric Signal Preprocessing:** High-frequency, synchronised neonatal vital sign data, including electrocardiograms (ECG), chest impedance (CI), and peripheral oxygen saturation (SpO_2) streams, are collected and preprocessed. To create distinct chronological data blocks, waveforms are systematically truncated into 30-second moving windows with a 5-second time shift [4, 9]. Python-based algorithms are used to clean signals, remove motion artefacts, and perform standard data normalisation [2, 6].
- **Phase 2: Predictive Machine Learning Modeling:** A pool of 47 distinct temporal and trend features is extracted from the preprocessed waveforms to capture cross-channel physiological correlations [7, 9]. To ensure clinical generalisability and prevent mathematical overfitting, models are evaluated using leave-one-out (LOPO) cross-validation. Deep neural networks (CNN and LSTM configurations) are trained alongside traditional machine learning architectures, such as Elastic Net Logistic Regression, Random Forest, and Support Vector Machines. To reduce false positives, a consecutive window detection criterion is implemented, which generates an alert only when an anomaly is verified across 5 consecutive window sets [4, 9].
- **Phase 3: Explainable AI (XAI) Architecture:** To increase clinician trust, SHapley Additive Explanations (SHAP) are used to calculate the precise quanti-

tative contribution of each vital sign feature to the model’s output [5]. Grad-CAM is integrated to generate visual activation overlays directly onto respiratory time-series, mapping algorithmic decisions against human expert clinical parameters [6].

- **Phase 4: Digital Hardware Translation:** The optimised network’s weights and biases are converted from signed floating points to 15-bit integers by multiplying each parameter by a scaling factor of 2^{10} . This innovative parameter reduction technique maps the compact model across a 260-cell layout within an Artix-7 FPGA environment, measuring physical circuit area and ensuring that the system remains within an efficient, low-power operational budget[8].

The selected analytical and windowing techniques are highly appropriate for this problem because they directly optimise event precision while bypassing the data buffering overheads of traditional sequential monitoring loops.

Chapter 7

Research Limitations

This study’s constraints are divided into clear design and methodological limitations, which influence its overall clinical application.

- **Design limitations and impacts:** The study focuses on a small, single-center retrospective dataset with geographically concentrated patient cohorts. This constraint directly limits the generalisability and external validity of the results across diverse intensive care settings. Additionally, the small training sample size prevents the deployment of much deeper deep learning layers due to an increased risk of mathematical overfitting[4, 6, 10].
- **Methodology limitations and impact:** The framework uses a retrospective, multi-stage windowing mechanism without real-time, online adaptive learning capabilities. This approach limits the system’s ability to adjust to live clinical shifts or concept drift, requiring manual retraining interventions when structural breaks occur in the time series. Furthermore, omitting direct respiratory flow measurements prevents the model from performing multiclass classification to distinguish between mixed or obstructive apneas and central events [4, 9, 11].

Chapter 8

Project Timeline

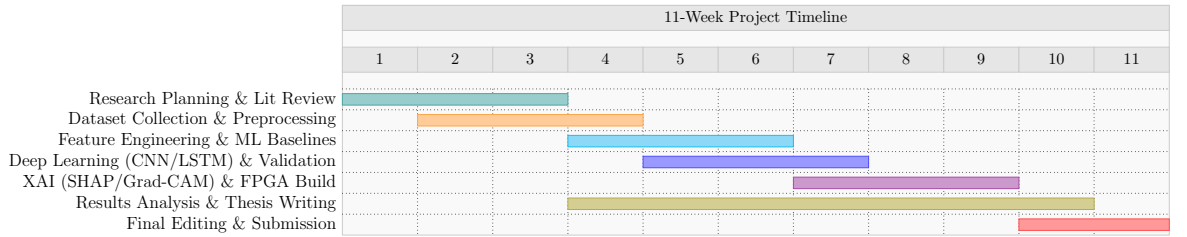


Figure 8.1: MSc Project 11-Week Horizontal Implementation Schedule.

To ensure a rigorous and structured approach to the project lifecycle, the proposed 11-week implementation plan is divided into seven sequential macro-phases that drive the research from its inception to final submission. The first month is spent laying the theoretical and data foundations, beginning with research planning and a thorough literature review, followed by dataset collection and continuous signal preprocessing. Raw high-frequency neonatal vital signs, such as electrocardiograms (ECG), chest impedance, and peripheral oxygen saturation (SpO_2) waveforms, will be cleaned, filtered of motion artefacts, and segmented into 30-second moving windows to prepare for robust algorithmic modelling. Weeks 4-7 are dedicated to feature engineering, baseline machine learning training, and deep learning optimisation with Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures evaluated using a leave one out cross-validation scheme. Once the core model's performance has been stabilised, explainable AI layers based on the SHAP and Grad-CAM frameworks will be integrated during Weeks 7 and 8, providing transparent, clinically auditable visual decision pathways. The final phase scales the trained network parameters to 15-bit integers for logic verification on an Artix-7 FPGA chip, followed by comprehensive results analysis, thesis compilation, and final submission by the August 28, 2026 deadline.

Chapter 9

Expected Findings

The expected findings of this study are entirely consistent with the stated research objectives and methodology:

- **Predictive Performance and Alarm Optimisation:** The study anticipates that integrating multi-parametric ECG, chest impedance, and SpO_2 waveforms into an optimised Convolutional Neural Network architecture will significantly outperform traditional threshold configurations. The CNN model is expected to achieve a central apnea detection precision of at least 0.54 and an overall sensitivity of 0.75, exceeding the conventional 0.35 precision floor . The system is expected to achieve high sensitivity for critical patient events, reaching 0.87 for apneas with bradycardia and 0.88 for apneas with low heart rates. These elevated metrics are expected to minimise clinical alert vulnerabilities by providing an automated 66% reduction in false monitoring warnings and up to an 80% reduction in acoustic alarm triggers via predictive subcritical windows [4, 7, 9].
- **Explainable AI and Interpretability Metrics:** The integration of SHAP and Grad-CAM layers is expected to yield clear visual explanation tools and transparent audit trails directly overlaying high-frequency time-series data. The model is expected to confirm that features derived from chest impedance provide the most relevant clinical contributions, with the highest median log odds ratios compared to other vital signs. Visualising these quantitative metrics allows care teams to observe the direct alignment between human expert reasoning and algorithmic decisions, establishing clinician trust for applications like discharge planning while not reducing model classification accuracy [5, 6, 9].
- **Digital Hardware Realisation Constraints:** Converting signed floating-

point parameters to 15-bit integers using a scaling factor of 210 is expected to result in a highly resource-efficient digital layout, in line with hardware translation objectives. The compiled feedforward design, configured across 260 cells on an Artix-7 FPGA, is expected to maintain an overall infant apnea detection accuracy rate above 85%. The typical power consumption budget is expected to be below 9W, indicating an optimised engineering footprint for future point-of-care deployment inside a standard neonatal nasal cannula [8].

References

- [1] S. B. Walker et al., “Novel approaches to capturing and using continuous cardiorespiratory physiological data in hospitalized children,” *Pediatric Research*, vol. 93, no. 2, pp. 396–404, Jan. 2023, doi: <https://doi.org/10.1038/s41390-022-02359-3>.
- [2] K. Shanmugapriya, S. Mangai, K. R. Laksita, S. Muthukumari, V. Sowmiya, and M. Subhasankari, “Smart Neonatal Supine Monitoring System with Vital Signs and Thermal Regulation,” *Advances in Engineering Research*, pp. 851–862, 2026, doi: https://doi.org/10.2991/978-94-6239-654-8_66.
- [3] C. Sitaula, E. Grooby, T. C. Kwok, D. Sharkey, F. Marzbanrad, and A. Malhotra, “Artificial intelligence-driven wearable technologies for neonatal cardiorespiratory monitoring. Part 2: artificial intelligence,” *Pediatric Research*, Dec. 2022, doi: <https://doi.org/10.1038/s41390-022-02417-w>.
- [4] G. Varisco, Z. Peng, H. Niemarkt, P. Andriessen, C. van Pul, and X. Long, “Improving central apnea detection in preterm infants: Neural network strategies and clinical impact,” *Computer Methods and Programs in Biomedicine*, vol. 272, p. 109038, Dec. 2025, doi: <https://doi.org/10.1016/j.cmpb.2025.109038>.
- [5] K. Helm et al., “Oxycardiorespirograms in neonatal monitoring—current gaps and future potential of artificial intelligence: a mini-review,” *Frontiers in Pediatrics*, vol. 13, Oct. 2025, doi: <https://doi.org/10.3389/fped.2025.1680074>.
- [6] A. Nikolaou et al., “Artificial Intelligence in Neonatal Respiratory Care: Current Applications and Future Directions,” *Applied Sciences*, vol. 16, no. 3, p. 1339, Jan. 2026, doi: <https://doi.org/10.3390/app16031339>.
- [7] R. Joshi, Z. Peng, X. Long, L. Feijs, P. Andriessen, and C. Van Pul, “Predictive Monitoring of Critical Cardiorespiratory Alarms in Neonates Under Intensive Care,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019, doi: <https://doi.org/10.1109/JTEHM.2019.2953520>.
- [8] O. Hassan, D. Parvin, and Syed Kamrul, “Machine Learning Model Based Digital Hardware System Design for Detection of Sleep Apnea Among Neonatal Infants,” pp. 607–610, Aug. 2020, doi: <https://doi.org/10.1109/mwscas48704.2020.9184554>.
- [9] G. Varisco et al., “Central apnea detection in premature infants using machine learning,” *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107155, Nov. 2022, doi: <https://doi.org/10.1016/j.cmpb.2022.107155>.
- [10] A. M. Millana et al., “A systematic review on the use of artificial intelligence in the neonatal intensive care unit: far beyond the potential impact,” *Seminars in Fetal and Neonatal Medicine*, vol. 31, no. 1, p. 101690, Nov. 2025, doi: <https://doi.org/10.1016/j.siny.2025.101690>.

- [11] L. Porcelli, M. Trovati, and F. Palmieri, “Real-time anomaly detection in seasonal time series with conditional variational autoencoder,” *Applied Soft Computing*, vol. 184, p. 113761, Aug. 2025, doi: <https://doi.org/10.1016/j.asoc.2025.113761>.