RADIANODE

Radianode Engineering Readiness Series

ERS-2026-001

Engineering Readiness Assessment

INT8 Post-Training Quantization for Seq2Point NILM on Resource-Constrained Edge Platforms

INDEPENDENT ENGINEERING VALIDATION

Reference	ERS-2026-001
Publication Version	1.1
Publication Date	July 2026
Engineering Domain	Edge AI · Embedded Systems · Intelligent Infrastructure
Status	Public
Prepared By	Radianode Engineering
Technical Review	Engineering Review Complete
Reproducibility	Supported
Supporting Engineering Artefacts	Official Radianode GitHub Repository

Executive Summary

This Engineering Readiness Assessment evaluates whether INT8 post-training quantization provides a deployment-ready optimisation for a Seq2Point Non-Intrusive Load Monitoring (NILM) model intended for resource-constrained Edge AI platforms.

The investigation was conducted using the Radianode engineering validation methodology, focusing on the engineering characteristics that determine deployment readiness rather than theoretical model performance.

The measured evidence demonstrates that the evaluated optimisation:

- Reduced deployment memory by 75%
- Preserved model behaviour
- Maintained prediction accuracy
- Introduced no observable deployment-critical numerical instability

Based on the engineering evidence presented in this assessment, the evaluated model satisfies the criteria for deployment readiness.

ENGINEERING DECISION

Can INT8 post-training quantization be adopted as a deployment optimisation for the evaluated Seq2Point NILM model without compromising engineering confidence?

This publication answers that question using measured engineering evidence.

INVESTIGATION SCOPE

Item	Description
Engineering Asset	Seq2Point NILM Model
Engineering Objective	Evaluate whether INT8 post-training quantization preserves deployment readiness of a Seq2Point NILM model for resource-constrained Edge AI platforms

Item	Description
Optimisation Evaluated	INT8 Post-Training Quantization
Deployment Scenario	Resource-Constrained Edge AI
Evaluation Dataset	UK-DALE
Model Architecture	Seq2Point
Implementation Framework	MATLAB Deep Learning Toolbox
Engineering Methodology	Radianode Engineering Methodology
Engineering Process	RVP (Radianode Validation Process)

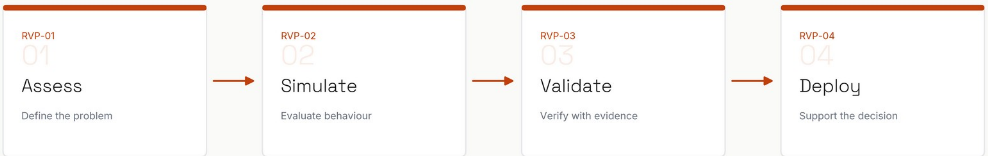
Engineering Process

The Radianode Validation Process (RVP™)

Every Engineering Readiness Assessment published by Radianode is executed using the Radianode Validation Process (RVP).

RVP provides a consistent engineering process for evaluating connected systems through independent assessment, engineering analysis, validation and evidence-based recommendations to reduce technical uncertainty before deployment.

THE FOUR STAGES OF RVP



RVP-01 — Assess Define the engineering problem by reviewing requirements, understanding system architecture, identifying constraints, and establishing the scope of assessment.

RVP-02 — Simulate Apply appropriate engineering analysis to evaluate expected system behaviour through simulation, modelling, hardware evaluation, RF analysis, digital twin engineering, or other suitable techniques.

RVP-03 — Validate Verify engineering assumptions using objective evidence, performance evaluation, documented findings, and structured engineering review.

RVP-04 — Deploy Provide the engineering evidence, readiness assessment, and independent technical recommendations required to support deployment decisions.

APPLICATION OF RVP WITHIN THIS ASSESSMENT

RVP Stage	Application in ERS-2026-001
Assess	Defined the engineering objective, deployment scenario, and evaluation scope for INT8 post-training quantization

RVP Stage	Application in ERS-2026-001
	of a Seq2Point NILM model.
Simulate	Performed engineering evaluation of the optimised model using representative residential energy data.
Validate	Assessed memory efficiency, behaviour preservation, prediction accuracy, and numerical stability through repeatable engineering workflows.
Deploy	Determined deployment readiness based on the measured engineering evidence presented in this assessment.

01

Memory Efficiency

ENGINEERING QUESTION

Has deployment memory been reduced sufficiently to provide a meaningful engineering benefit?

Model	Deployment Memory
FP32	354.10 KB
INT8	88.52 KB
Measured Reduction	75%

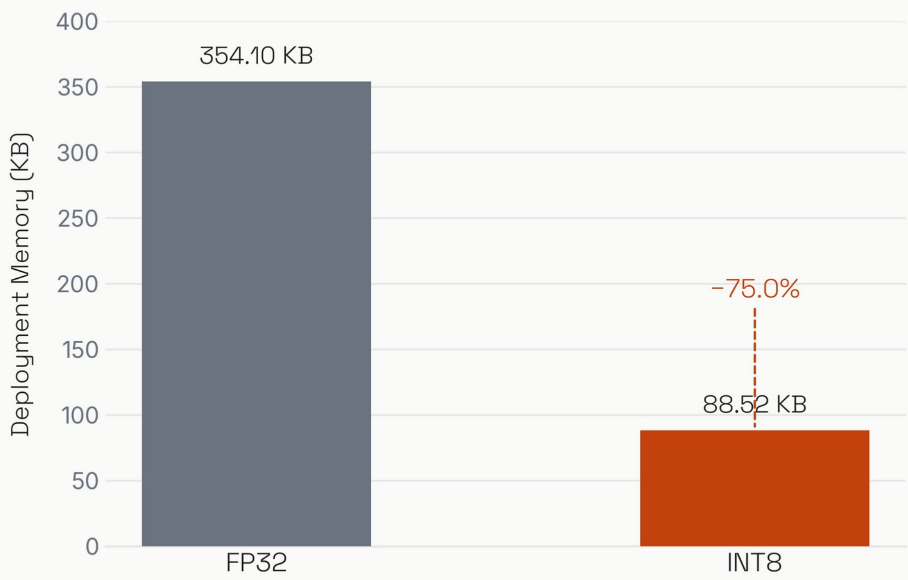


FIGURE 1 *Memory Footprint Comparison*

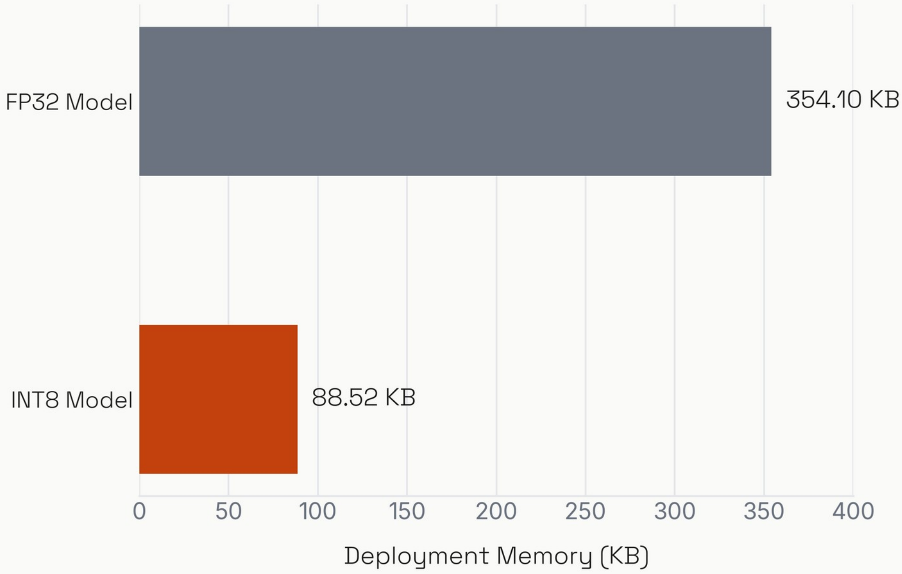


FIGURE 2 *Model Memory Breakdown*

> **ENGINEERING INTERPRETATION**
>
> The optimisation reduced deployment memory by 75% without modifying the network architecture or requiring model retraining. This represents a significant improvement in deployment efficiency and increases suitability for memory-constrained embedded platforms.

✓ **Engineering Finding —** Memory efficiency improved with no observed engineering compromise.

02

Behaviour Preservation

ENGINEERING QUESTION

Did the optimised model continue behaving like the original implementation?

Metric	Result
Correlation	0.9940
Mean Difference	0.8532 W
Maximum Difference	2.7062 W

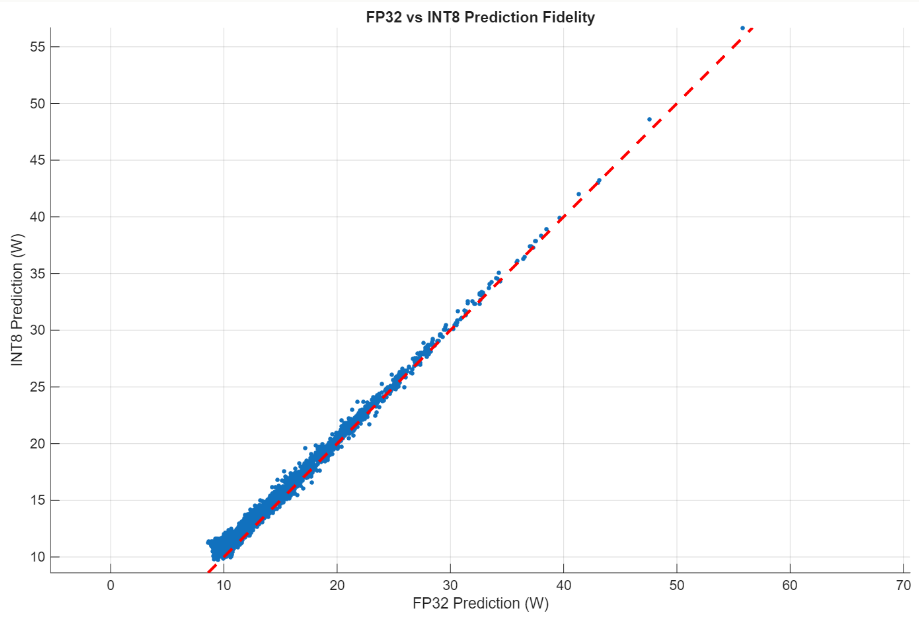


FIGURE 3 *FP32 vs INT8 Prediction Scatter*

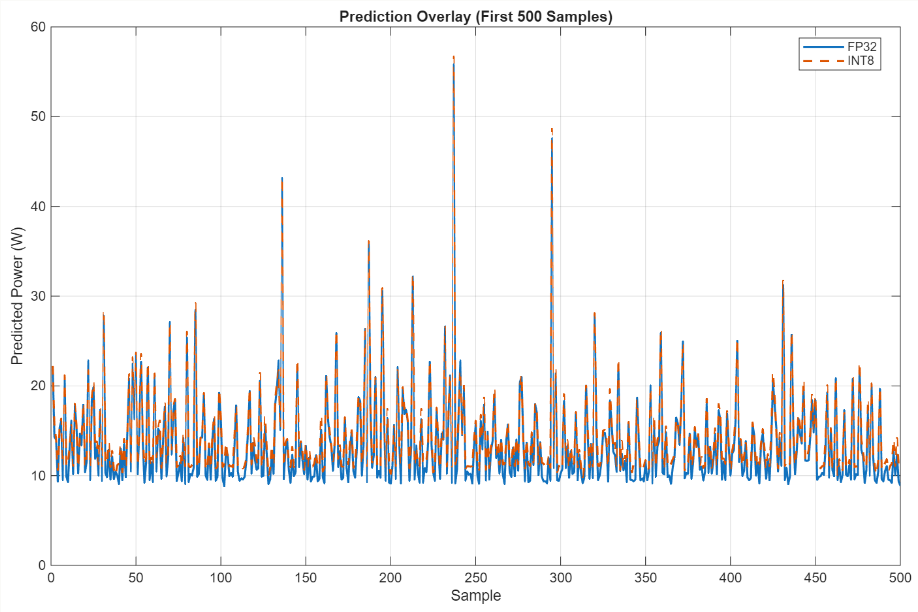


FIGURE 4 *Prediction Overlay*

> **ENGINEERING INTERPRETATION**
>
> Prediction behaviour remained highly consistent following optimisation. The measured correlation demonstrates that the optimised model preserved the operational characteristics of the original implementation.

✔ **Engineering Finding —** Model behaviour remained consistent after optimisation.

03

Prediction Accuracy

ENGINEERING QUESTION

Was inference performance maintained following optimisation?

Metric	FP32	INT8
RMSE	21.6061 W	21.4706 W
MAE	11.7867 W	11.5046 W

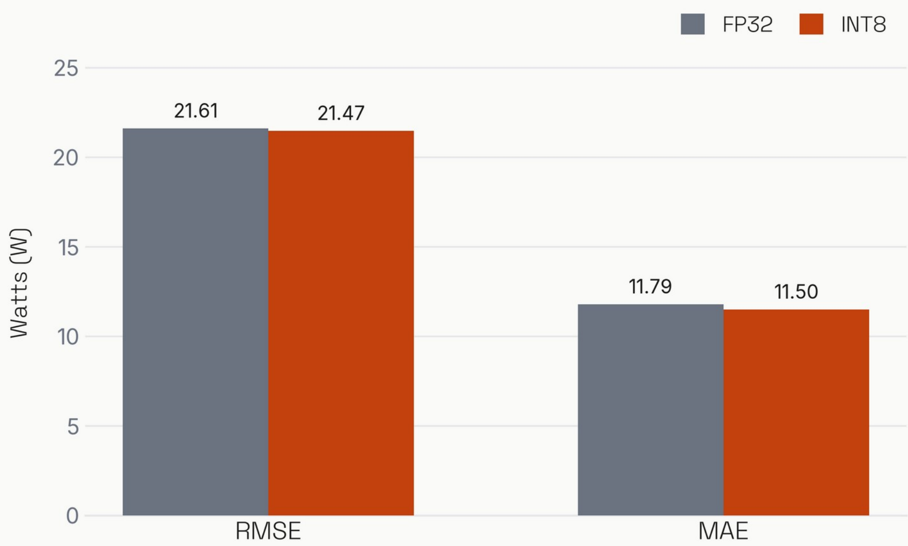


FIGURE 5 *Accuracy Comparison*

> **ENGINEERING INTERPRETATION**
>
> Prediction accuracy remained effectively unchanged. The measured differences are operationally insignificant and provide no evidence that optimisation reduced inference performance.

✓ **Engineering Finding —** Prediction accuracy was maintained after optimisation.

04

Numerical Stability

ENGINEERING QUESTION

Did optimisation introduce deployment-critical numerical effects?

Metric	Result
FP32–INT8 RMSE	1.0254 W
FP32–INT8 MAE	0.8532 W
Mean Relative Error	7.90%

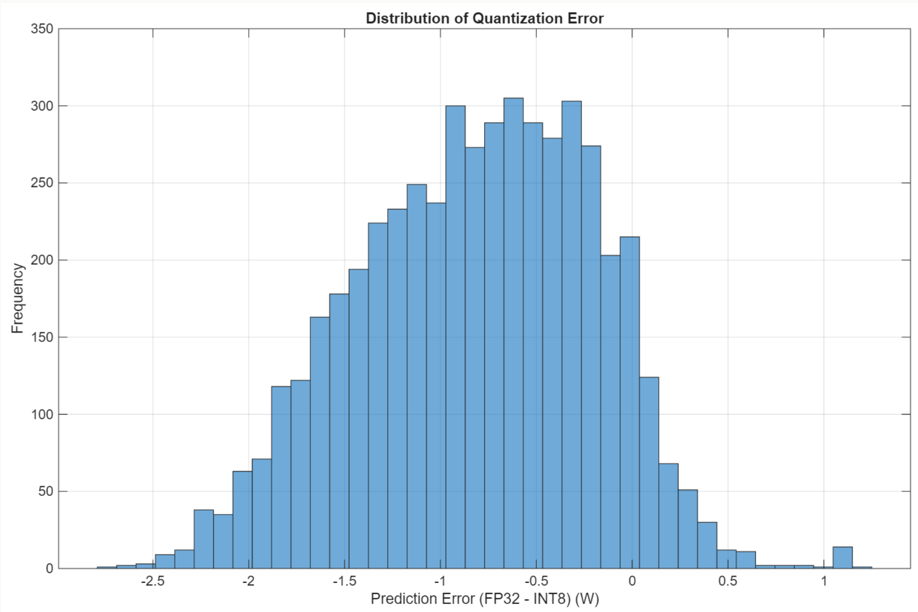


FIGURE 6 *Error Distribution Histogram*

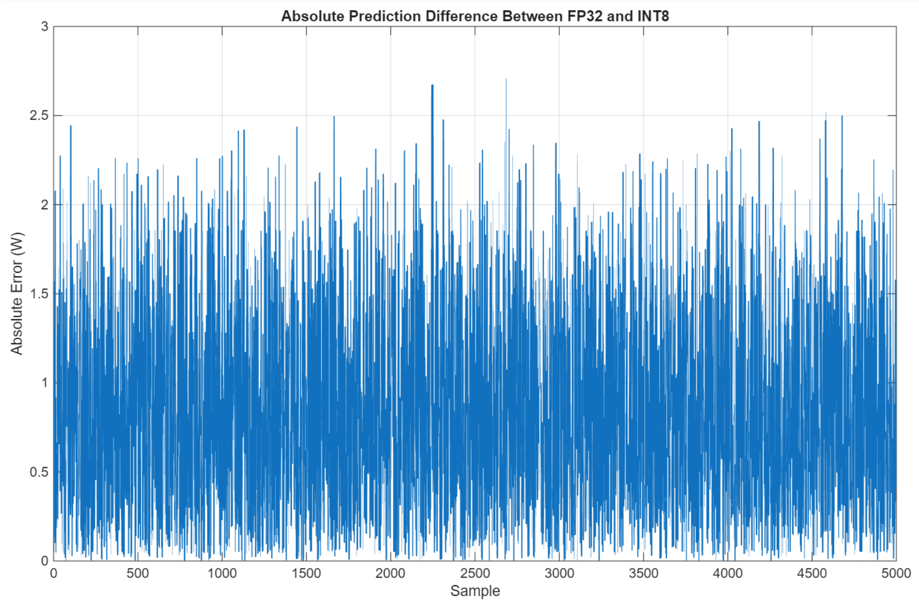


FIGURE 7 *Absolute Error*

ENGINEERING INTERPRETATION

The observed numerical differences remained small, predictable and non-systematic throughout the evaluation.

The measured relative error did not introduce observable changes in prediction behaviour or inference quality.

Within the scope of this engineering assessment, the measured numerical deviation provides no evidence of deployment-critical numerical instability.

✓ **Engineering Finding —** Numerical stability remained within acceptable engineering limits.

Scope of Evaluation

Clearly defining the limits of this assessment strengthens the credibility of the engineering conclusions presented.

THIS ASSESSMENT EVALUATES

- ✓ Deployment memory efficiency
- ✓ Behaviour preservation
- ✓ Prediction accuracy
- ✓ Numerical stability

THIS ASSESSMENT DOES NOT EVALUATE

- — Hardware execution latency
- — Energy consumption
- — Processor utilisation
- — Hardware-specific optimisation
- — Cross-dataset generalisation
- — Hardware-in-the-loop deployment
- — Long-term operational performance

ENGINEERING EVIDENCE SUMMARY

Engineering Criterion	Outcome
Memory Efficiency	✓ Confirmed
Behaviour Preservation	✓ Confirmed
Prediction Accuracy	✓ Maintained
Numerical Stability	✓ Verified

Radianode Engineering Position

Following independent engineering evaluation, Radianode considers INT8 post-training quantization to be an appropriate deployment optimisation for the evaluated Seq2Point NILM model when deployment objectives prioritise memory efficiency without compromising inference performance.

For the evaluated deployment scenario:

- Deployment memory was significantly reduced.
- Prediction behaviour remained consistent.
- Inference accuracy was maintained.
- Numerical stability remained within acceptable engineering limits.

Based on the measured evidence, no engineering barriers to deployment were identified.

Deployment validation on the intended production hardware remains recommended as part of final system acceptance.

ENGINEERING READINESS OUTCOME

✓ READY FOR EDGE DEPLOYMENT

Reproducibility Statement

This Engineering Readiness Assessment is an implementation of the Radianode Validation Process (RVP).

Supporting engineering artefacts, reproducibility resources, implementation workflows, and publication assets are available through the official Radianode GitHub repository.

REPOSITORY CONTENTS

- Engineering implementation
- Evaluation workflows
- Figure-generation scripts
- Engineering metrics
- Publication assets
- Reproducibility guide
- Version history

REPOSITORY

github.com/radianode/ers-2026-001

APPENDIX A

Figure Traceability

Every published figure is traceable to the engineering workflow that generated it.

Report Figure	Engineering Workflow	Output
Figure 1 – Memory Footprint Comparison	Memory Evaluation	Memory comparison chart
Figure 2 – Model Memory Breakdown	Memory Evaluation	Deployment memory chart
Figure 3 – FP32 vs INT8 Prediction Scatter	Behaviour Comparison	Prediction fidelity scatter plot
Figure 4 – Prediction Overlay	Behaviour Comparison	Prediction overlay plot
Figure 5 – Accuracy Comparison	Accuracy Evaluation	RMSE/MAE comparison chart
Figure 6 – Error Distribution Histogram	Numerical Stability Evaluation	Error distribution histogram
Figure 7 – Absolute Prediction Error	Numerical Stability Evaluation	Absolute prediction error plot

This appendix strengthens reproducibility without exposing or depending on a specific implementation toolchain.

Continue the Conversation

Every connected system presents different engineering constraints.

The question is rarely whether a technology can work. The real question is whether it is ready for your deployment.

At Radianode, we help engineering teams answer that question through independent validation, simulation-driven engineering and vendor-neutral technical assessment.

Whether you're evaluating Edge AI, wireless communications, embedded hardware, intelligent infrastructure or complex connected systems, we can help you make engineering decisions with greater confidence.

OUR ENGINEERING CAPABILITIES

— Engineering Readiness Assessments
— Hardware Performance Validation
— Edge AI Validation & Optimisation
— Wireless Systems Engineering
— Digital Twin Engineering
— Vendor-Neutral Technical Advisory

SUPPORTING ENGINEERING ARTEFACTS

The engineering evidence presented in this publication is supported by a reproducibility package available through the official Radianode GitHub repository. The repository contains the engineering implementation, evaluation workflows, engineering metrics, publication assets, and supporting documentation required to reproduce the evidence presented in this assessment.

github.com/radianode/ers-2026-001

If you're facing an engineering decision and need objective technical evidence before deployment, we'd welcome the opportunity to discuss your project.

TALK TO AN ENGINEER

hello@radianode.com
www.radianode.com