Edge Radar Material Classification Under Geometry Shifts

Jannik Hohmann, Dong Wang, and Andreas Nüchter

Abstract— **Material awareness can improve robotic navigation and interaction, particularly in conditions where cameras and LiDAR degrade. We present a lightweight mmWave radar material classification pipeline designed for ultra-low-power edge devices (TI IWRL6432), using compact range-bin intensity descriptors and a Multilayer Perceptron (MLP) for real-time inference. While the classifier reaches a macro-F1 of 94.2% under the nominal training geometry, we observe a pronounced performance drop under realistic geometry shifts, including sensor height changes and small tilt angles. These perturbations induce systematic intensity scaling and angle-dependent radar cross section (RCS) effects, pushing features out of distribution and reducing macro-F1 to around 68.5%. We analyze these failure modes and outline practical directions for improving robustness with normalization, geometry augmentation, and motion-aware features.**

I. INTRODUCTION

Robots deployed for long periods in the wild must maintain reliable perception under changing environments and gradual sensor drift. Material classification is a deployment-sensitive capability: it can inform traversability and interaction decisions, yet sensing geometry (height, tilt, and mounting tolerances) inevitably varies over time. Radar is attractive for such settings because it is largely independent of ambient illumination and directly measures electromagnetic reflections rather than visual appearance [1].

Practical radar material classification on embedded devices faces two coupled challenges. First, edge hardware imposes tight compute and power budgets, favoring compact feature representations and small networks [2]. Second, intensity-based radar signatures are strongly geometry-dependent: received power varies with range and changes with angle-dependent RCS, so small mounting variations can yield out-of-distribution (OOD) inputs at deployment.

This paper reports a baseline implementation and an evaluation centered on geometry shifts. We implement a lightweight pipeline on the TI IWRL6432 platform that extracts a 12-dimensional range-bin intensity descriptor and performs real-time inference with an MLP on-device. Under the nominal geometry, the model achieves a macro-F1 of 94.2%. However, performance degrades substantially under realistic geometric perturbations (height and tilt), revealing a robustness gap. We analyze the dominant failure modes and provide a roadmap toward more geometry-aware and deployment-robust radar material classification that remains compatible with embedded constraints. We emphasize that the goal of this study is not to claim geometry-invariant material classification, but to establish an edge-deployable intensity-based baseline and to quantify where and why such a baseline fails under realistic deployment shifts.

This paper makes the following contributions:

- We present a practical edge baseline for radar material classification on TI IWRL6432, combining a compact intensity descriptor with an MLP for real-time on-device inference.
- We quantify robustness under geometry shifts (height and tilt) and under session-level variations using macro-F1 and confidence statistics.
- We identify failure modes caused by range-dependent intensity scaling and angle-dependent RCS changes, and outline mitigation strategies (normalization, augmentation, and motion-aware features).

II. RELATED WORK

FMCW mmWave radar has become a relevant modality for robotic perception because it measures electromagnetic reflections and can operate independently of ambient illumination [1]. Recent work emphasizes resource-efficient on-device processing, enabling "machine learning on the edge" for low-power radar platforms [2] [3] [4]. Such constraints often motivate compact feature representations and lightweight models (e.g., small MLPs) when memory and latency budgets are strict [5].

Radar-based material (or surface) classification leverages differences in reflectivity, dielectric properties, and scattering behavior, but signatures can be sensitive to geometry, multipath, and sensor-dependent factors, especially when relying on absolute intensity cues [6] [7]. Beyond intensity-only descriptors, learning-based approaches increasingly use richer inputs such as IQ signals or radar images. For example, SM-CNet demonstrates supervised surface material classification using mmWave radar IQ signals with complex-valued CNNs and explicitly studies distance generalization [8]. Material-ID explores mmWave-based material identification with broader system considerations and data-driven modeling [9]. Earlier work such as RaCaNet shows that radar imaging with deep CNNs can achieve high classification accuracy in controlled settings for 60 GHz radar material classification [10].

Related RF sensing literature (outside strict mmWave radar pipelines) further highlights that material recognition can be feasible from RF propagation effects, but also prone to distribution shifts. RadarCat explores radar categorization for

This work was in parts supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag under the grant number KK5150106RL4.

Jannik Hohmann, Dong Wang (dong.wang@uni-wuerzburg.de) and Andreas Nüchter are with the Department of Informatics XVII (Robotics), Julius-Maximilians-University Würzburg, 97074 Würzburg, Germany. Andreas Nüchter is also with the Zentrum für Telematik e.V., Würzburg and currently International Visiting Chair at U2IS, ENSTA, Institut Polytechnique de Paris, France.

interactive sensing, illustrating the utility of radar signatures for recognition tasks under practical constraints [11]. Wi-Fi based material sensing similarly studies how channel responses can encode material properties, emphasizing feasibility but also sensitivity to environment and setup [12]. In industrially motivated settings, 60 GHz radar has also been explored for material sorting/classification in recycling-oriented scenarios, showing strong accuracy under constrained setups [13]. More recently, feature-selection-based ML pipelines have been investigated for 60 GHz FMCW radar material classification, again underlining that the feature/geometry interplay is critical for performance [14].

OOD robustness and long-term reliability are widely recognized as central challenges for deployment-grade perception [15] [16]. In addition, run-time monitoring is increasingly viewed as a practical safety layer to detect or anticipate failures of learning-based perception components during operation [17]. Motivated by these insights, our work quantifies how small, realistic geometric changes (height and tilt) can break an edge-deployed intensity-based mmWave material classifier, and highlights mitigation directions compatible with embedded constraints.

III. METHODOLOGY

The pipeline follows a deterministic signal-processing chain designed for high-frequency real-time execution. Raw radar measurements are transformed into compact feature vectors, which are then classified by a lightweight neural network.

A. Data Acquisition and Hardware Configuration

Data collection was performed using the Texas Instruments IWRL6432BOOST, an integrated single-chip FMCW mmWave radar sensor operating in the 60 GHz band [18]. The sensor was mounted in a nadir-looking geometry at a constant height of $H = 45\,\text{cm}$ with a nominal downward inclination of $\theta = 90^\circ$ relative to planar material samples. We considered five material classes: iron, aluminum, plexiglass, wood, and limestone.

The nominal dataset was recorded under the fixed reference geometry shown in Fig. 1. For each material class, the same acquisition procedure and frame-integration setting were used. The resulting sequences were split into training and validation subsets with the same acquisition geometry and class set. Unless stated otherwise, all reported validation results refer to data not used during training but recorded under this nominal geometry.

To reduce stochastic temporal noise, intensity measurements were integrated over a fixed number of frames per recorded sequence, using the identical setting across all classes and experiments. The robustness datasets modify one factor relative to the reference setup: sensor height, inclination angle, or recording session/sample identity. The independent-recording dataset was acquired again at $H = 45\,\text{cm}$ and $\theta = 90^\circ$, but in separate recording sessions using independently positioned material samples. This setting is intended to separate pure geometry shifts from session-level changes such as sample placement, small environmental variations, and measurement noise.

In-plane rotations of the planar material samples were not systematically varied in this study. This is a limitation of the current evaluation and is left for future work together with broader target geometries and outdoor measurements.

Fig. 1 illustrates the default acquisition geometry used as the in-distribution reference domain (*Known Data*).

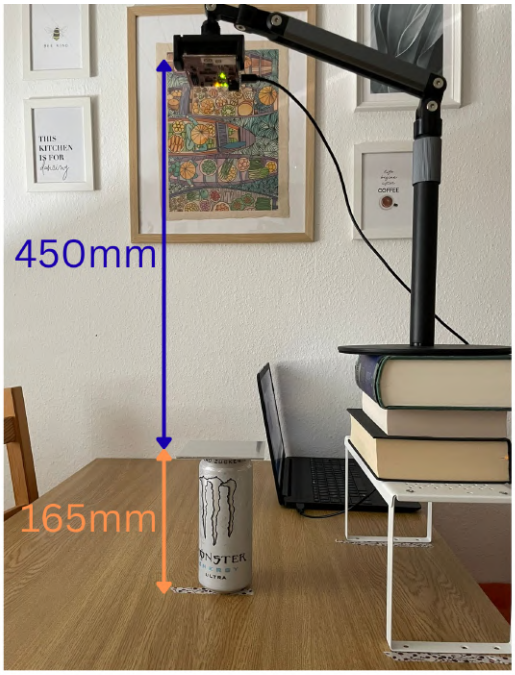

Fig. 1: Default experimental setup used for baseline data acquisition. The radar is mounted in a nadir-facing configuration at $H = 45\,\text{cm}$ and $\theta = 90^\circ$ relative to the material surface.

B. Feature Extraction

The extraction process converts raw Analog-to-Digital Converter (ADC) data into range-intensity profiles. To comply with edge constraints, we use a low-dimensional descriptor based on range-dependent magnitudes instead of full range-Doppler tensors. We deliberately do not apply range- or angle-dependent intensity normalization in this baseline. This design keeps the descriptor simple for edge deployment, but also makes the model sensitive to geometry-induced amplitude changes, which is the main robustness limitation analyzed in Sec. IV.

1) Range Bin Selection: We determine an informative range-bin window by analyzing energy concentration on the training set and selecting the bins that consistently cover the dominant return from the planar target at nominal geometry. In our setup, relevant material-specific information concentrates within Range Bins 6–17. Bins outside this interval predominantly contain background clutter or noise-

floor components and are excluded to reduce compute and improve stability.

2) Intensity Descriptor: For each frame, the magnitudes of the 12 selected range bins are concatenated into a 12-dimensional feature vector. This descriptor acts as direct input to the classifier and captures material-dependent backscattering characteristics shaped by dielectric properties and scattering coefficients [6].

C. Model Architecture and Edge Deployment

We implement a compact Multilayer Perceptron (MLP) to satisfy latency and memory constraints on the IWRL6432. The network consists of:

- **Input layer:** 12 nodes corresponding to the selected range bins.
- **Hidden layers:** two fully connected layers with 32 and 16 neurons, respectively, each followed by Batch Normalization [19] and ReLU activations [20].
- **Output layer:** 5 nodes with Softmax to obtain class probabilities.

The model is trained using the Adam optimizer with cross-entropy loss. The training hyper-parameters were kept fixed for all reported experiments. Model selection is performed on the nominal-geometry validation split. No geometry-shifted data are used during training, so the height, tilt, and independent-recording experiments evaluate out-of-distribution generalization.

After training, the network is deployed on-device via the IWRL6432 software stack. The MLP forward pass achieves an average on-device inference latency of approximately 20 μs, enabling execution within the sensor processing cycle without external compute.

IV. EXPERIMENTAL EVALUATION

We first quantify performance under nominal (training) geometry and then probe generalization under realistic geometric perturbations. We report macro-averaged F1-score and analyze Softmax confidence distributions. Table I summarizes the evaluated data splits and geometry conditions. The nominal validation split is used to quantify in-distribution performance, whereas all other splits are used only for robustness evaluation.

TABLE I: Summary of evaluated data splits and geometry conditions.

Split	Height	Tilt	Purpose
Nominal validation	45 cm	90°	In-distribution validation
Height shift	35 cm	90°	Range robustness
Height shift	55 cm	90°	Range robustness
Tilt shift	45 cm	80° / 100°	Angle robustness
Independent recordings	45 cm	90°	Session/sample robustness

A. Baseline Performance under Nominal Geometry

The baseline evaluation uses the nominal configuration ($H = 45\,\text{cm}$, $\theta = 90°$).

1) Classification Accuracy and Confusion Analysis: The classifier reaches a macro-F1 score of 94.2% on the validation split. The confusion matrix in Fig. 2 shows strong separability of metallic vs. non-metallic classes. Residual confusion is mainly between iron and aluminum, which exhibit similar reflectivity in the 60 GHz band.

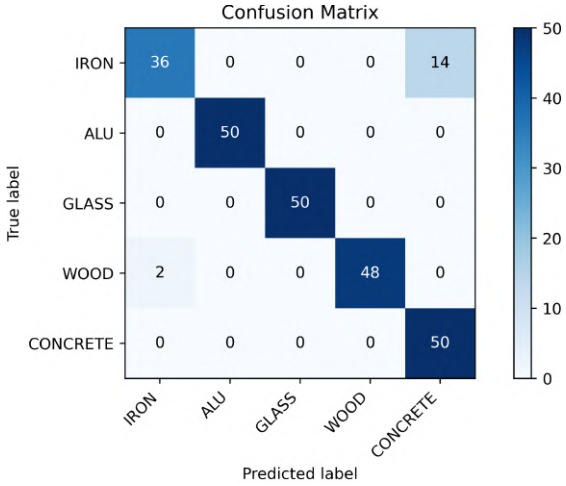

Fig. 2: Confusion matrix of the MLP classifier under nominal conditions ($H = 45\,\text{cm}$, $\theta = 90°$). Diagonal entries indicate per-class recall.

2) Confidence Distribution: For the nominal dataset, correct predictions typically yield high Softmax maxima. Fig. 3 shows that confidence values for correct classifications often exceed $p = 0.9$, while misclassifications correlate with lower confidence, consistent with proximity to learned decision boundaries.

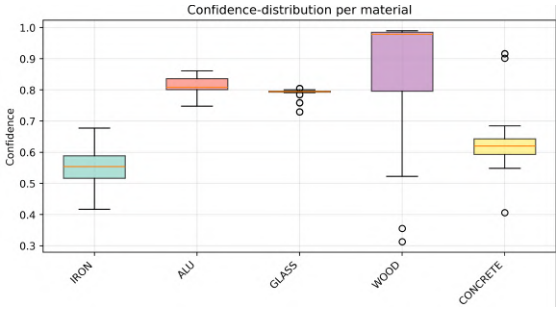

Fig. 3: Confidence (maximum Softmax probability) distribution under nominal geometry.

B. Performance under Geometry Shifts

We evaluate two geometry perturbations relative to the nominal setup: height change ($H \in \{35, 55\}\,\text{cm}$) and tilt change ($\theta = 90° \pm 10°$).

1) Impact of Sensor Height Displacement: We vary sensor height between 35 cm and 55 cm. Performance degrades as range deviates from the training height, consistent with the strong range dependence of received power. Because the MLP operates on absolute intensities within fixed range bins, feature vectors drift relative to learned decision boundaries.

Fig. 4 shows increased off-diagonal mass, particularly when increased distance reduces the apparent intensity of reflective materials.

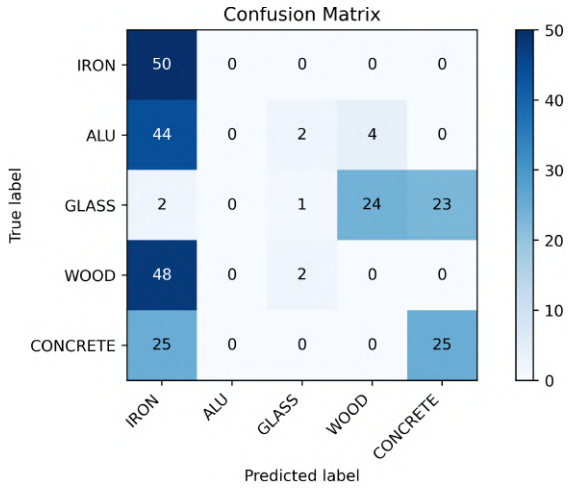

Fig. 4: Confusion matrix under height variations (35 cm and 55 cm). Off-diagonal density increases due to distance-dependent power changes.

2) Impact of Inclination Angle and RCS: We tilt the sensor by $\pm 10^{\circ}$ from nadir. This perturbs the effective RCS of planar targets; for specular reflectors (iron, aluminum), small angular changes can redirect energy away from the monostatic receiver, reducing measured intensity. We observe a macro-F1 decrease to approximately 68.5%. The confidence distribution in Fig. 5 shifts toward lower values, indicating elevated uncertainty under OOD geometry.

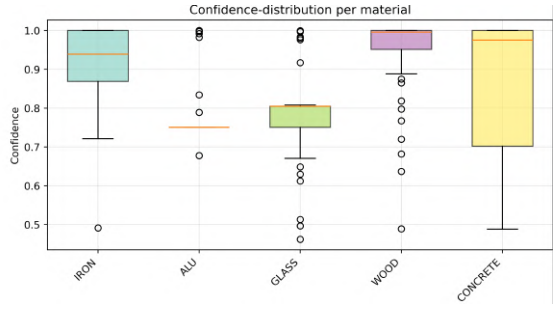

Fig. 5: Confidence distribution under geometry shifts (tilt). Increased low-confidence predictions indicate reduced certainty under OOD geometry.

C. Performance under Similar Geometry with Independent Recordings

To separate geometry shift from session-level variability, we evaluate a "similar but unknown" dataset recorded under the same nominal geometry ($H = 45\,\text{cm}$, $\theta = 90^{\circ}$), but in independent recording sessions. The material classes are unchanged, but the samples are repositioned and recorded separately from the training data. Thus, this experiment captures session-level effects such as small placement differences, environmental changes, SNR fluctuations, and sample-instance variability, while keeping the nominal height and tilt fixed.

1) Generalization to Material Classes: Compared to the baseline validation, we observe a notable performance drop and increased confusion (Fig. 6). This suggests sensitivity to session-specific intensity fluctuations and partial overfitting to the training recordings' noise characteristics, particularly for low-contrast material pairs.

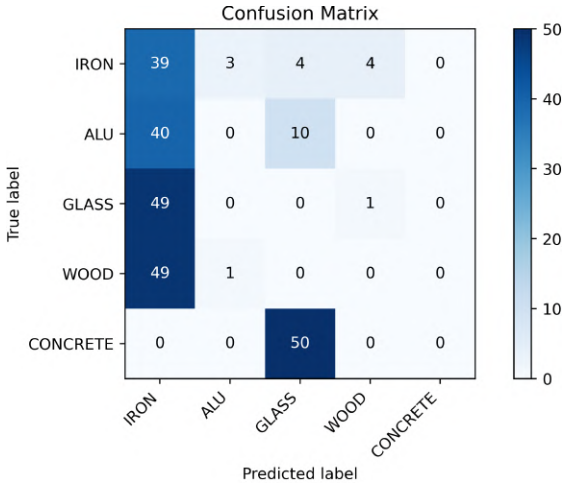

Fig. 6: Confusion matrix for "similar but unknown" data at nominal geometry ($H = 45\,\text{cm}$, $\theta = 90^{\circ}$). Increased confusion indicates session-level sensitivity despite unchanged geometry.

2) Comparative Confidence Analysis: Fig. 7 shows a shift toward lower confidence compared to the baseline, consistent with feature vectors moving closer to decision boundaries due to session-level intensity changes and SNR fluctuations.

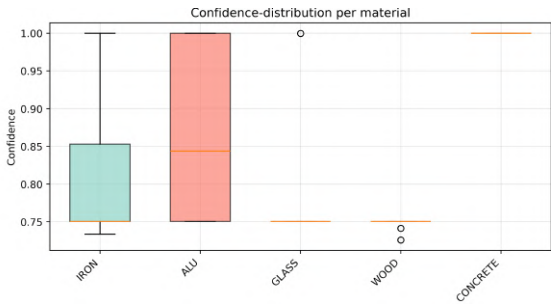

Fig. 7: Confidence distribution for "similar but unknown" data at nominal geometry. The probability mass shifts toward lower confidence compared to the baseline.

V. CONCLUSION AND DISCUSSION

We presented a lightweight mmWave radar material classification pipeline for ultra-low-power edge deployment on the TI IWRL6432BOOST [18]. Using a 12-dimensional range-bin intensity descriptor and a compact MLP, the system achieves strong in-distribution performance (macro-F1 94.2%) with inference latency around 20 μs, making it suitable for embedded operation [2].

A. Failure Modes under Geometry Shifts

A central finding is that absolute intensity features are not geometry-invariant. Height changes induce systematic scaling of received power, while small tilt angles alter effective RCS and can strongly attenuate specular returns [7]. Both effects yield OOD feature drift and large performance degradation (macro-F1 down to $\sim$68.5% under tilt). The experiments do not fully disentangle the effect of model capacity from the effect of the feature representation. However, the consistent degradation under range and tilt changes indicates that a major source of error is the geometry dependence of the absolute range-bin intensities. A larger model alone would not remove this dependence unless the training data or preprocessing also covers the corresponding amplitude and RCS variations. Therefore, we interpret the observed failures as a combined limitation of the compact descriptor and the training distribution, rather than as an isolated limitation of the MLP architecture. Furthermore, even under unchanged geometry, independent recordings exhibit session-level variability that reduces confidence and accuracy, indicating sensitivity to temporal noise patterns.

B. Roadmap Toward Robust Deployment

Our results suggest three practical directions that remain compatible with embedded constraints:

- **Range-aware normalization:** compensate distance-dependent scaling (e.g., calibration curves or physics-guided normalization).
- **Geometry augmentation:** train with height/tilt perturbations to enlarge the support of the training distribution.
- **Motion-aware features:** explore Doppler or micro-motion signatures to reduce reliance on absolute intensity cues [21].

In addition, simple uncertainty monitoring using confidence statistics can provide a lightweight trigger for re-calibration or fallback behaviors in long-term operation.

C. Future Work

Beyond normalization and augmentation, future work should investigate lightweight architectures with improved inductive bias (e.g., small CNNs) while respecting memory and latency constraints. Disciplined hyper-parameter selection (learning rate schedules, weight decay) may further stabilize training across recording sessions [22].

a) Limitations: Our study focuses on planar samples, one radar platform, and a limited set of geometry perturbations. In-plane rotations, larger tilt and height ranges, additional material instances, outdoor conditions, and alternative feature representations are left for future work. Moreover, the current results characterize an intentionally compact intensity-based edge baseline; they should not be interpreted as the upper bound of radar material classification performance.

REFERENCES

[1] S. Rao, "Introduction to mmwave Sensing: FMCW Radars," in *Texas Instruments Technical Article SWRA553A*, March 2017.

[2] M. Yanik, A. K. Chandrasekaran, and S. Rao, "Machine Learning on the Edge with the mmWave Radar Device IWRL6432," in *Texas Instruments Technical Article SWRA781*, May 2023.

[3] Texas Instruments, *IWRL6432 Single-Chip 57- to 64GHz Industrial Radar Sensor Datasheet (Rev. B)*, 2025, sWRS298B, revised March 2025. [Online]. Available: https://www.ti.com/lit/ds/symlink/iwrl6432.pdf

[4] ——, "IWRL6432BOOST Evaluation Board (BoosterPack) Product Page," Online, accessed 2026-03-05. [Online]. Available: https://www.ti.com/tool/IWRL6432BOOST

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[6] R. Weber, *Physik /1: Klassische Physik: experimentelle und theoretische Grundlagen*, 1st ed. Wiesbaden: Teubner, 2007.

[7] C. Wolf, "Radar Signature," https://www.radartutorial.eu/01.basics/rb66.en.html, n.d., accessed: 2025-06-13.

[8] S. Hägele, F. Seguel, D. Salihu, A. Misik, and E. Steinbach, "Smcnet: Supervised surface material classification using mmwave radar iq signals and complex-valued cnns," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2025.

[9] G. Chen, C. Luo, H. Zeng, G. Wen, Z. Luo, J. Wang, J. Zhang, Z. Yang, and J. Li, "Material-id: Towards mmwave-based material identification," *ACM Transactions on Sensor Networks*, vol. 21, no. 4, pp. 41:1–41:26, 2025. [Online]. Available: https://doi.org/10.1145/3742432

[10] J. Weiß and A. Santra, "Material classification using 60-ghz radar and deep convolutional neural network," in *2019 International Radar Conference (RADAR)*, 2019, pp. 1–6.

[11] H. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley, "Radarcat: Radar categorization for input & interaction," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST 2016)*. ACM, 2016, pp. 833–841. [Online]. Available: https://doi.org/10.1145/2984511.2984515

[12] D. Zhang, J. Wang, J. Jang, J. Zhang, and S. Kumar, "On the feasibility of wi-fi based material sensing," in *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom 2019)*. ACM, 2019, pp. 41:1–41:16. [Online]. Available: https://doi.org/10.1145/3300061.3345442

[13] T. Albing and R. Nelander, "Material classification of recyclable containers using 60 ghz radar," arXiv preprint, 2023. [Online]. Available: https://arxiv.org/abs/2312.14539

[14] S. Nongpromma and N. Shutimarrungson, "An investigation of material classification using 60 ghz mmwave fmcw radar and feature selection-based machine learning," in *2025 22nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2025.

[15] A. Schwaiger, P. Sinhamahapatra, J. Gansloser, and K. Roscher, "Is Uncertainty Quantification in Deep Learning Sufficient for Out-of-Distribution Detection?" in *Proceedings of the 1st International Workshop on Artificial Intelligence Safety Engineering (WAISE 2020)*, 2020.

[16] M. Henne, J. Ganslöser, A. Schwaiger, and G. Weiss, "Machine Learning Methods for Enhanced Reliable Perception of Autonomous Systems," Fraunhofer IKS, Tech. Rep., 2021, whitepaper.

[17] Q. M. Rahman, P. Corke, and F. Dayoub, "Run-time monitoring of machine learning for robotic perception: A survey of emerging trends," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/abs/2101.01364

[18] Texas Instruments, "IWRL6432BOOST EVM User Guide," 2022. [Online]. Available: https://www.ti.com/lit/ug/swru596/swru596.pdf

[19] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv preprint arXiv:1502.03167*, 2015. [Online]. Available: https://arxiv.org/abs/1502.03167

[20] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010, pp. 807–814.

[21] C. Wolf, "Doppler-Effekt," https://www.radartutorial.eu/11.coherent/co06.de.html, n.d., accessed: 2025-06-08.

[22] L. N. Smith, "A disciplined approach to neural network hyper-parameters," arXiv:1803.09820, 2018.