(Received 25 September 2024, Accepted 4 November 2024, Scheduled 12 November 2024)

Refinement of Chipless RFID Tags across Multiple Positions for Improved Recognition Reliability through Machine Learning Techniques

Athul Thomas*, Midhun M. Sylaja, and James Kurian

Cochin University of Science and Technology, Kochi 682022, Kerala, India

ABSTRACT: Chipless Radio Frequency Identification (RFID) technology offers a cost-effective and durable alternative to chipped tags for identification and tracking applications. By eliminating the need for an integrated circuit, chipless tags are cheaper and can withstand harsher environments. This opens doors to not only track items throughout a supply chain or monitor valuable assets, but also integrate basic sensors for functionalities like environmental monitoring or smart agriculture. However, limitations in data capacity, read range, and decoding complexity currently hinder their full potential. This paper explores the application of machine learning techniques to improve the interrogation process and enhance the reliability of chipless RFID systems. The effectiveness of machine learning in optimising chipless RFID systems hinges on the richness and variety of training data. A robust dataset encompassing diverse tag characteristics, environmental factors, and reader configurations is paramount. Nevertheless, gathering real-world RFID data can be difficult. To address this, a data collection procedure has been specifically designed to gather backscattered information from the chipless tags at multiple orientations and distances. Four binary combinations of a 5-bit RFID tag based on frequency-selective surfaces operating in the 2–8 GHz range are considered for generating the database. The dataset is then used to train and validate various classification models, including support vector machine (SVM), k-nearest neighbour (k-NN), Decision Tree (DT), Naive Bayes classifier, and Logistic Regression (LR). The proposed Support Vector Machine model is able to identify the tag at a distance of up to 70 cm from the interrogator, with multiple rotational degrees of freedom.

1. INTRODUCTION

Radio Frequency Identification (RFID) technology provides significant advantages in the realm of the Internet of Things (IoT) by enabling efficient and automated identification and tracking of objects. This leads to improved inventory management, streamlined logistics, and enhanced data collection [1–4]. The main setback to its adoption is the cost of designing and fabricating the integrated circuit required for data processing. Chipless RFID (CRFID), introduced in recent years, offers the aforementioned advantages while being affordable, printable, and suitable for large-scale manufacturing and hostile settings [5].

In a conventional RFID framework, the interrogator antenna initiates the activation of RFID elements and subsequently captures the scattered electromagnetic signals. Each resonant element on the tag modulates the incident wave with a unique resonant frequency signature determined by its design parameters. Chipped RFID systems leverage an application-specific integrated circuit (ASIC) to enhance the backscattered signal, enabling efficient separation of tag data from ambient noise. In contrast, CRFID tags rely solely on their Radar Cross Section (RCS), leading to a restricted reading range and increased susceptibility to interference and reflections. The time and frequency domain characteristics of chipless backscattered tags, essential for data encoding and reader interaction, are analysed in [6] by Babaeian and Karmakar. The intricate nature of the reflected signal in CRFID systems, a result of factors like antenna reflection, tag structural modes, tag antenna modes, environmental reflections, and background noise, is mathematically described by Equation (1). These components can overlap significantly in both the time and frequency domains, creating complex signals that are difficult to decipher.

$$S_{total} = S_{antenna} + S_{tag\_struc} + S_{tag\_ant} + S_{env} + noise \quad (1)$$

where S_{total} is the total signal, and $S_{antenna}$, S_{tag_struc}, S_{tag_ant}, and S_{env} are the antenna reflection component, tag structural-mode components, tag antenna-mode components, and the reflection from the environment, respectively.

The system's performance is further influenced by factors such as the distance and orientation between the interrogator and tag, and the presence of nearby objects. To address these hurdles, numerous efforts have been made focusing on tag design, encoding techniques, and detection algorithms. A comprehensive overview of measurement methods, response detection approaches, and decoding techniques is provided by Brinker and Zoughi [7].

Despite the progress in tag design and detection algorithms, ensuring reliable tag identification continues to be a hurdle. Several factors influence success, including the tag's efficiency, the effectiveness of the chosen signal processing techniques,

* Corresponding author: Athul Thomas (athulthomas@cusat.ac.in).

TABLE 1. Review of existing literature.

Reference	Calibration	Data set size	Model	Achievements	Limitations
[21]	Background subtraction	Exact dataset size is not mentioned	SVM, k-NN and Ensembles	Compared different ML models on the collected dataset Not sensitive to tag misprints	Size of the dataset is not mentioned
[20]	Background subtraction	d — 5 cm to 50 cm, Multiple orientations (0° to 40° along ϕ, 0° to 40° along θ) (Total 816 measurements)	SVM	Achieved high accuracy along multiple orientation Presence of nearby objects considered	Very small dataset
[22]	Time gating	10 meas/tag (180 instances) & data augmentation (3600 instances)	DNN	18 different tags considered	Artificial dataset Measurement inside anechoic chamber
[23]	Background subtraction	Three different datasets (8800 instances)	LR	Introduction of dimensionality reduction method, thresholding scheme	Generalisation of the model needs background - subtraction
This Paper	Raw S_{21} data is used	Two different datasets (4480 instances /tag)	SVM	Proposed data collection method Considered multiple orientations Effect of background noise is considered	Model fails beyond certain orientation and distance

and the surrounding environment. The existing methods often require sophisticated algorithms, each specifically designed to decode the unique properties and encoding schemes of a particular tag [8–10]. This approach, while functional, presents constraints in terms of scalability and adaptability. A paradigm shift is emerging in chipless radio frequency identification with the arrival of machine learning (ML) algorithms. These algorithms represent a significant leap forward, moving away from the traditional methods that require specialised knowledge about tag design and intricate signal processing techniques [11-14]. Instead, data-driven algorithms harness the power of data itself. By analysing vast datasets collected from CRFID interrogations, these algorithms can learn to identify patterns and decode the information embedded within the complex reflected signals.

This study addresses two primary concerns that impede the use of ML for CRFID: data scarcity and algorithm robustness. The issue of limited availability of training data is tackled by automating the data acquisition system and streamlining the process to collect a large volume of data that accurately reflects real-world scenarios. By replacing manual collection, the data's relevance to practical deployments is ensured, ultimately enhancing the performance of the trained ML algorithms. Additionally, the research investigates the robustness of these ML algorithms under various real-world conditions, evaluating their capability in recognising tags across long detection ranges and different tag orientations.

2. STATE OF THE ART ML IN CRFID

In conventional tag identification and authentication systems, effectively extracting tag information requires expertise in advanced signal processing techniques, including background subtraction, time gating, continuous wavelet transformation, and match filtering [7, 15–19]. Recent studies indicate that ML methods could address the difficulties related to CRFID tag identification [20–23]. However, there is currently limited literature on the application of ML for RFID tag classification. Table 1 provides an overview of some existing contributions in this area.

A hybrid method combining millimetre-wave frequency scanning and ML approaches to enhance the performance of CRFID systems is reported by Arjomandi and Karmakar [21]. Individual tags printed with alphanumeric characters are interrogated using orthogonally polarised horn antennas in the 60 GHz spectrum. The resulting data undergoes processing for feature learning and classification using various networks such as SVM, k-NN, and ensembles. This approach achieves a

reading accuracy of over 95%, marking significant progress in low-cost chipless tag applications.

An ML-based tag identification method capable of effectively reading transponders across different ranges and situations is proposed by Jeong et al. [20]. The dataset includes 816 measurements from four RFID topologies with T-shaped resonant elements operating in the 2–10 GHz bandwidth. The dataset encompasses the magnitude, phase, real part and imaginary part of the reflection coefficient S_{21}, enhancing the comprehensiveness of the database. Classification techniques such as DTs, Boosted Trees, k-NN, and SVM are employed, with the SVM classifier using a linear kernel achieving an accuracy of 99.3%.

The feasibility of using deep neural networks to analyse electromagnetic signatures from CRFID tags for authentication applications is investigated by Nastasiu et al. [22]. The study considered 18 distinct mm-wave E-shape resonators (65–72 GHz) in ten different locations, yielding a dataset comprising 180 measurements. Data augmentation techniques, including adding white Gaussian noise, were used to expand this dataset to 3600 observations, enhancing the diversity of information in each electromagnetic signature. A fully connected neural network evaluated on this augmented database successfully categorised 18 tags with 100% accuracy.

A comprehensive methodology for implementing ML algorithms to achieve reliable CRFID tag identification is provided by Sokoudjou et al. [23]. The study examines three distinct datasets covering various ranges, with and without initial background subtraction. Principal Component Analysis (PCA) is utilised for dimensionality reduction, alongside a thresholding scheme to boost prediction confidence. Results demonstrate accurate classification of analysed tags, achieving comparable accuracy values to those reported in existing literature.

The literature reveals that models are generally trained with relatively small datasets, often comprising fewer than 600 measurements per tag, which may constrain their generalisability to new tag responses. Automated data collection methods, typically burdensome in RFID measurements, are not widely covered. Most datasets are either augmented from limited real-life data or refined through background subtraction. Furthermore, evaluating the trade-offs between tag orientation and interrogator position remains necessary.

3. DATA COLLECTION

The efficacy of ML algorithms in optimizing CRFID systems is intrinsically tied to the richness, heterogeneity, and comprehensive representation of key parameters within the training dataset. These parameters include the angular orientations of the tag — specifically, the yaw (ψ), pitch (θ), and roll (ϕ) — along with the spatial separation (d) between the tag and interrogator, and the varying environmental conditions in which the system operates. A dataset that captures these multidimensional variables is essential for enabling ML models to generalize effectively across diverse scenarios, thereby enhancing the system's robustness and performance in practical applications. Given the challenges associated with acquiring real-world RFID data, particularly from chipless tags, a rigorously designed data collection protocol is required. This protocol must systematically capture the backscattered electromagnetic responses under varying angular orientations, distances, and environmental conditions, thus providing a scientifically sound basis for training ML models to address the current limitations in CRFID technology.

3.1. Tag Design and Topology

In this study, the proof-of-concept CRFID tag described by Betancourt et al. [24] serves as a reference. The octagonal tag is designed to operate in the ultra-wideband 2–8 GHz bandwidth. The symmetry provided by the octagonal shape enables the tag to be read independently of the polarisation of the interrogator. This topology is advantageous in real-life applications where maintaining a stable tag position is challenging. Fig. 1 illustrates the dimensions of a unit element that encodes 5 bits of information. Fig. 2 shows the simulated radar cross section results for the four tags analysed in this study, generated using CST Microwave Studio.

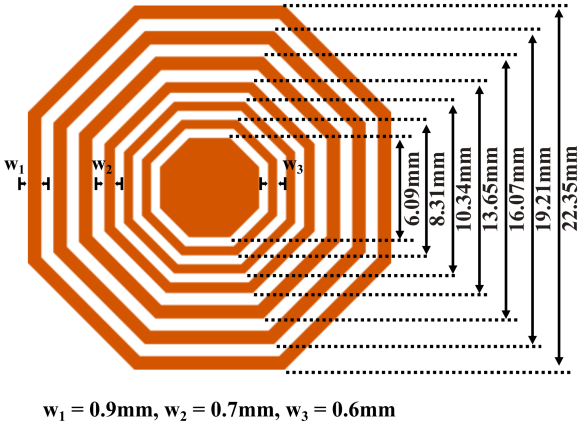

FIGURE 1. Detailed structure of the octagonal CRFID tag unit element highlighting key dimensions necessary for its design and functionality.

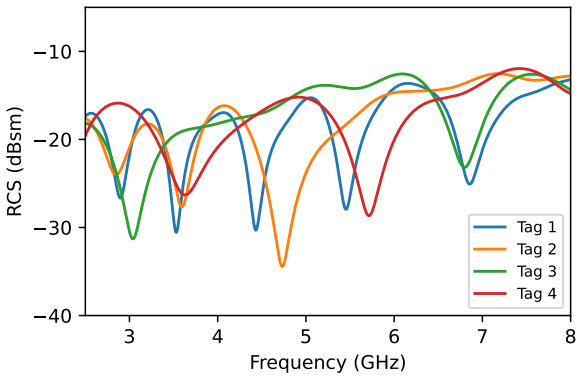

FIGURE 2. Simulated radar cross section results for the four tags analysed in this study, generated using CST Microwave Studio.

The distribution of surface current (j_{in} and j_{out}) in the adjacent octagons determines the presence or absence of response in the tag. When the adjacent tags are open-circuited, the cre-

ated electric field will be similarly polarised, and at bore sight, the scattered beam will destructively interfere, resulting in a response at a frequency determined by the diameter, width, and separation of the rings. Conversely, adding a short circuit between adjacent octagons causes the radiated electric field from the current sources to constructively interfere, suppressing resonance.

The tag design is based on frequency-selective surface (FSS) design methodology. Once a unit element, as shown in Fig. 1, is selected, this structure can be further replicated to enhance the overall Radar Cross Section. For the study reported here, an array of 2×2 unit elements, each with dimensions of 50 mm $\times$ 50 mm, is utilised. Fig. 3 depicts the RFID tag array fabricated on an FR-4 substrate.

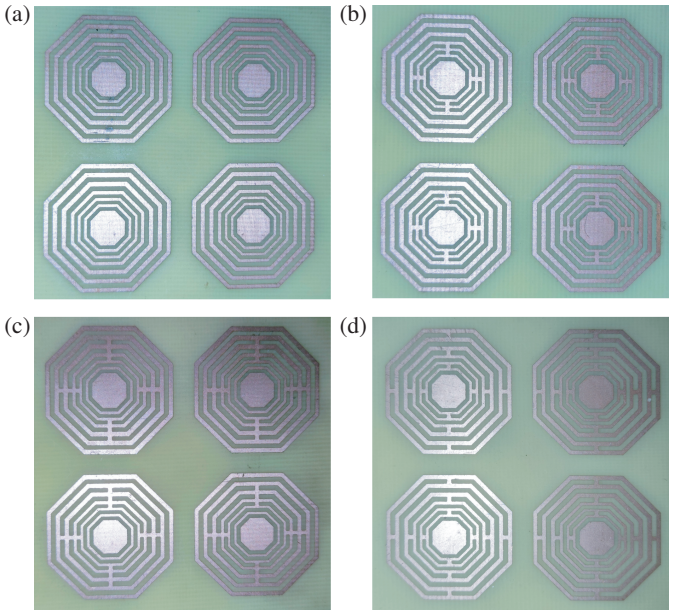

FIGURE 3. The tag fabricated on FR4 material as a 2×2 unit elements array. (a) Represents the 1-1111 encoded tag: All five octagonal rings contribute to resonance, representing the binary code 1-1111. (b) Represents the 1-0001 encoded tag: Short-circuits are added between adjacent rings to suppress the response, with the remaining rings contributing to resonance. (c) Tag corresponds to 1-0001 binary combination. (d) Tag corresponds to 0-1010 binary combination.

For the Radar Cross Section to be sufficiently diverse, the bit sequence in the RFID tags is chosen to have a maximum Hamming distance (dH) between them. Table 2 shows the bit combinations selected and the Hamming distance between individual tags. The tags exhibit a Hamming distance between 2 and 4.

TABLE 2. Hamming distance (dH) between selected Tags.

Tag	0-1010	1-0001	1-1100	1-1111
0-1010	0	4	3	3
1-0001	4	0	3	3
1-1100	3	3	0	2
1-1111	3	3	2	0

3.2. Experimental Setup

The block diagram of the tag positioning system is shown in Fig. 4, and the realisation of the actual experimental setup is showcased in Fig. 5. The measurement setup involves a tag holder, an interrogator, a controller, and a computer running a Python program. The tag holder is made of Acrylic sheet and a light weight dielectric foam and is attached to two servo motors at the bottom. These motors enable precise control over the tag's roll (ϕ) and yaw (ψ) orientations. The motors are strategically positioned away from the tag to eliminate potential reflections from their coils or shafts. A motor connected to the tag holder's axis directly controls the yaw movement, while a belt-driven mechanism facilitates roll motion. Pitch angle (θ) adjustments are achieved by inserting precisely angled dielectric foam cutouts between the tag and its holder.

A Rohde & Schwarz ZVB 20 Vector Network Analyser (VNA) [25] is employed as a 2–8 GHz microwave source, with two wide-band horn antennas operating in the 2–18 GHz frequency band serving as the interrogators. To ensure seamless data acquisition, a Python script interfaces the network analyser and servo motor controller. Communication between the VNA and PC is established via socket communication through port 5020. This setup allows the user to initialise various network analyser parameters by exchanging SCPI (Standard Commands for Programmable Instruments) commands. A micro-controller board supervising the servo motors communicates via UART, facilitating precise positioning of the tag for each measurement. The temporal lag associated with manoeuvring the servo motor to position the tag and configuring the analyser for executing the frequency sweep has been minimised to less than two seconds. This optimisation aims to enhance time efficiency between successive measurements.

The network analyser is configured with 201 data points per measurement, resulting in 201 distinct reflection values for the scattering parameter S_{21}, with intervals of 0.0348 GHz in the 2–8 GHz frequency span. To compensate for the potential limitations of the tag's size, which could lead to received backscattered power falling below the receiver's sensitivity threshold, the VNA's transmitted power is configured slightly higher at 5 dBm. Finally, to simulate a realistic operational environment and capture data that reflects real-world scenarios, the entire setup is positioned outside the confines of an anechoic chamber.

Determining the response of a CRFID tag (σ^{tag}) involves a detailed procedure, as described by Betancourt et al. [24]. The Network Analyzer is configured with proper time gating to isolate the reflections specifically from the RFID tag, eliminating other noise sources. The reflection coefficient (S_{21}) of the tag is then measured, indicating the amount of incident electromagnetic wave reflected by the tag. Subsequently, the Radar Cross Section is computed by subtracting the background noise from the measured reflection coefficient and normalising it using a known reference RCS as in Equation (2) [24], ensuring that the computed RCS is accurate and comparable across different measurements.

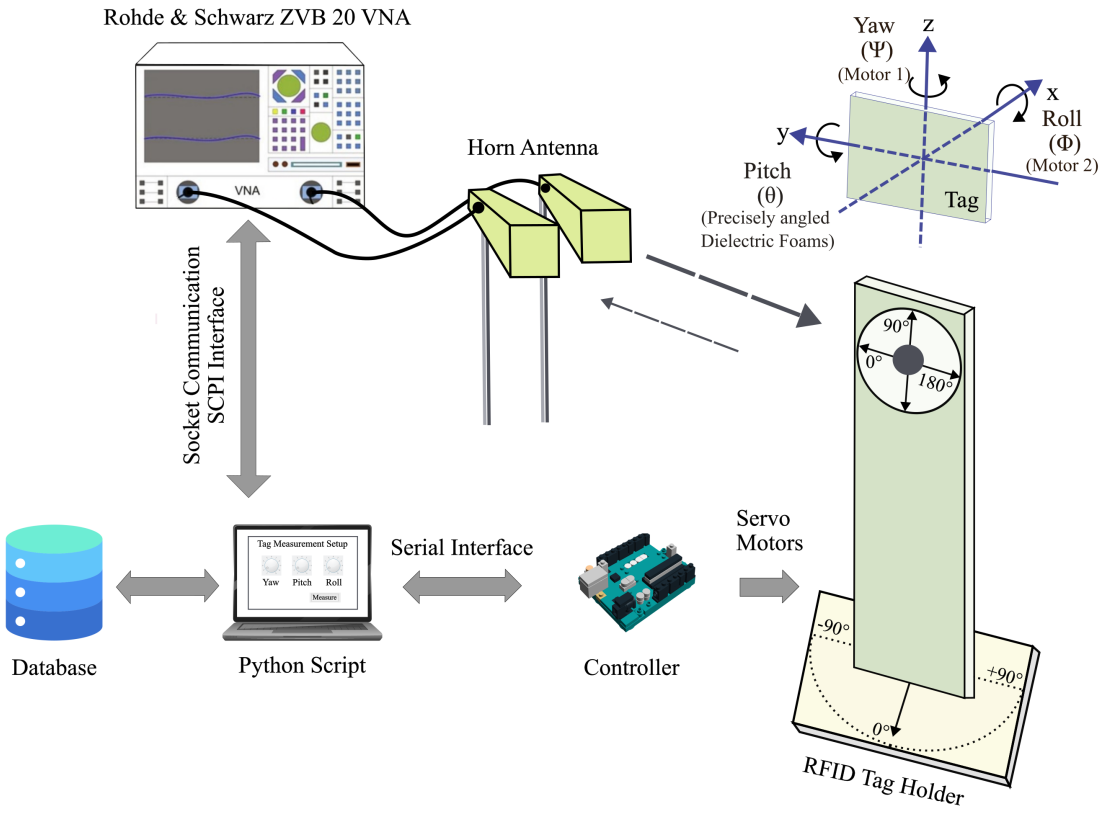

FIGURE 4. Block diagram illustrating the RFID data acquisition setup. The Python script serves as an intermediary, enabling communication between the network analyser and the servo motor via SCPI and Serial interfaces, respectively.

FIGURE 5. Experimental setup in a microwave lab for collecting RFID backscattered signals at varying distances and orientations.

$$\sigma^{tag} = \left[\frac{S_{21}^{tag} - S_{21}^{no\text{-}tag}}{S_{21}^{ref} - S_{21}^{no\text{-}tag}}\right]^2 \sigma^{ref} \tag{2}$$

where $S_{21}^{no\text{-}tag}$ is related to the measurement performed without tag. The S_{21}^{ref} represents a reference value obtained from a known measurement setup, and σ^{ref} corresponds to the simulation results obtained for this reference.

The RCS values computed through this approach at different positions are shown in Fig. 6(a) through Fig. 6(e). The response for each bit level is distinguishable across most positions and orientations. As the distance or orientation of the tag changes, the response is also affected, impacting the system's performance, as shown in Fig. 6(b) through Fig. 6(e).

In practical scenarios, estimating the Radar Cross Section by extracting background information and applying time-gating techniques may not be feasible due to operational constraints. One significant challenge is the dynamic nature of environments such as warehouses or production facilities, where continuous movement of personnel, machinery, and goods creates variability that complicates the establishment of a stable background reference for RCS estimation. Additionally, many CRFID applications require real-time data processing to make timely decisions, and the time needed for comprehensive background extraction and time-gating implementation often does not align with these urgent operational demands, leading to potential delays. Furthermore, the limitations of portable or handheld CRFID devices, which may lack the computational power necessary for complex algorithms, affect the ability to carry out accurate background noise subtraction and precise time gating.

Figure 6(f) displays the raw S_{21} data considered without applying time gating or background noise subtraction. In this raw form, the CRFID tag's responses are not distinctly identifiable. Despite this, a comparison with a metal plate reveals differences in the tag's responses, indicating that some identi-

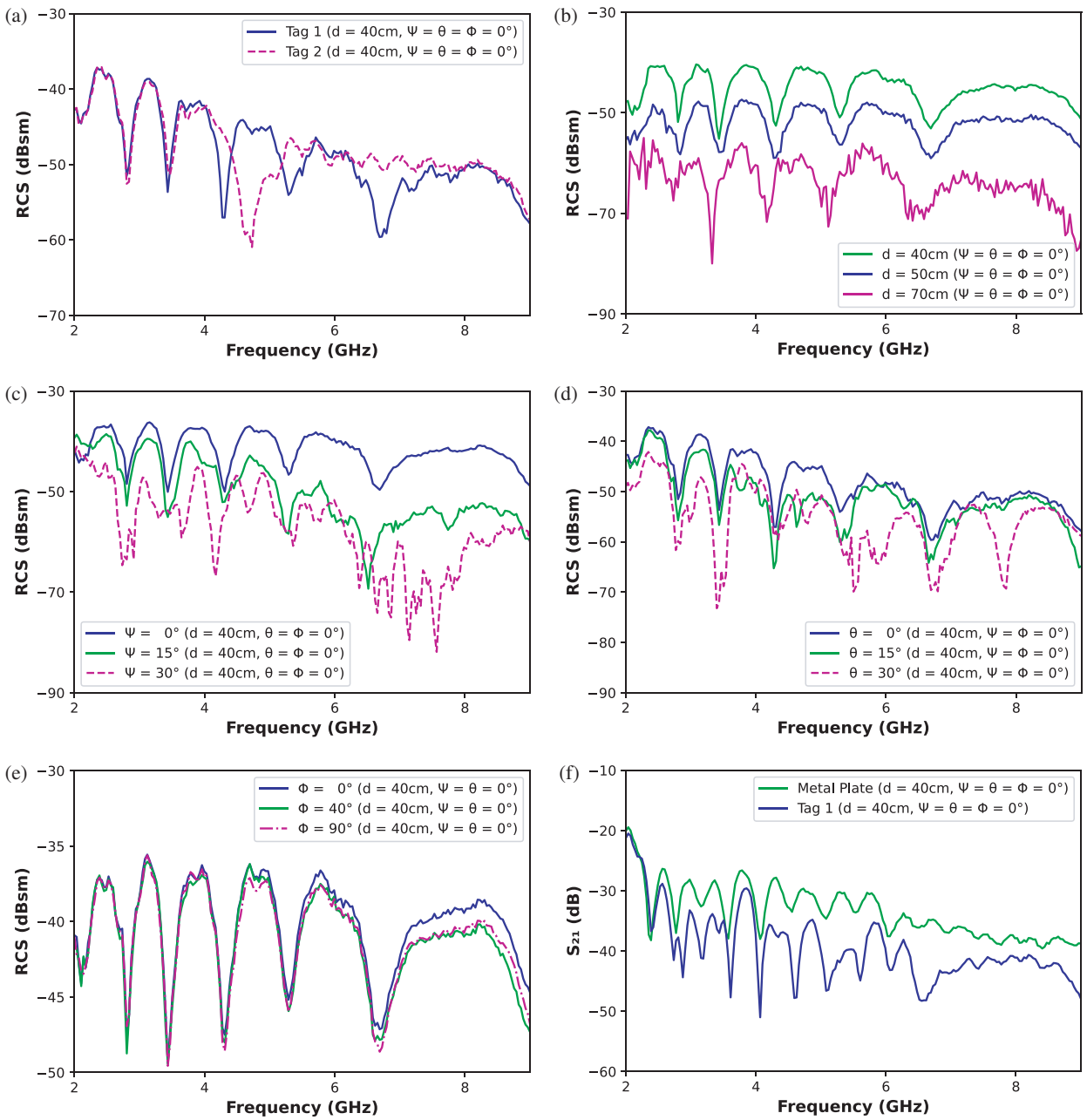

FIGURE 6. Response characteristics of RFID Tag 1 under various conditions: (a) Comparison of responses for two tags under identical conditions, (b) RCS at varying distances, (c) RCS at different yaw angles (ψ), (d) RCS at different pitch angles (θ), (e) RCS at different roll angles (ϕ), (f) Raw S_{21} compared with a metal plate..

fiable characteristics are still present. Identifying the underlying pattern with traditional approaches remains difficult due to overlapping signals and noise that obscure distinct responses. To overcome this, ML techniques can be applied to decode the tag data effectively. These techniques can discern patterns and correlations not easily detectable with conventional methods. However, for ML to be successful in this context, a sufficient amount of data is necessary. To address this, a methodology as shown in Fig. 7, focuses on creating a rich dataset that captures the key variables affecting RFID measurements. By carefully controlling various elements, the data acquisition methodology ensures the creation of a comprehensive and informative dataset that encompasses the diverse factors influencing RFID measurements.

At each position, S_{21} measurements are taken both with (S_{21}^{tag}) and without ($S_{21}^{no\text{-}tag}$) the tag to confirm the influence of external noises or variations in tag holder placements. The dataset includes a total number of 4,480 measurements per tag as shown in Table 3 with the following variables; 1) distance (d) from 40 cm to 100 cm in steps of 10 cm, 2) roll (ϕ) from 0° to 180° in steps of 20° 3) pitch (θ) from 0° to 70° in steps of 10° and 4) yaw (ψ) from 0° to 70° in steps of 10°.

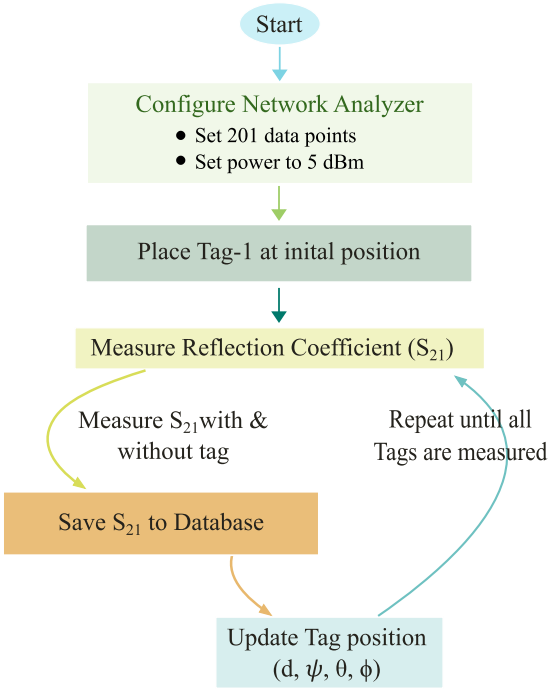

FIGURE 7. Flowchart outlining the procedural steps involved in conducting measurements for database generation, from initial setup to data recording and analysis.

TABLE 3. The range of parameters (distance, yaw, pitch, roll) considered for preparing the dataset for training the model.

Variable	Range	Number of measurements/tag
Distance (d)	40 cm to 100 cm in steps of 10 cm	7
Roll (φ)	0° to 180° in steps of 20°	10
Pitch (θ)	0° to +70° in steps of 10°	8
Yaw (Ψ)	0° to +70° in steps of 10°	8
Total		4480

4. ML METHODS

RFID tag recognition is a multi-class classification problem, with the primary goal of categorising tags into predetermined classes using distinctive characteristics retrieved from backscattered data. The nature of S_{21} data, which typically involves high-dimensional feature vectors representing the magnitude values of reflected signals, aligns well with the capabilities of ML algorithms. In the realm of RFID tag recognition, researchers have explored both traditional ML techniques and deep neural network (DNN) models. However, when being confronted with comparatively small datasets, ML methods have demonstrated more reliable classification performance. This preference stems from the inherent risks associated with employing deep learning approaches on limited data volumes, where over-fitting becomes a pronounced concern, potentially compromising the model's ability to generalise effectively [26, 27]. Traditional ML techniques, such as SVM, Random Forest (RF), and k-NN, have proven robust and effective in handling small datasets [28–30]. These algorithms are less susceptible to over-fitting, as they inherently prioritise simplicity and regularisation.

Figure 8 shows the proposed machine learning classification stages, illustrating each phase from data preprocessing to model evaluation and deployment. To select the best model which can effectively distinguish the tag under various conditions, an AutoML framework is employed [31]. These frameworks offer a systematic approach to model selection, hyper-parameter tuning, and model evaluation, making the process more efficient and accurate. As mentioned earlier, background noise data, in the absence of the original tag, is also included in the database for each RFID tag position. This enables more robust training of each model, enabling them to learn from both scenarios: when background noise is present and when it has been removed. The AutoML framework includes a variety of models such as SVM, k-NN, DT, LR, and Naive Bayes. It is supplied with two distinct datasets: one where background noise has been subtracted and the other containing raw S_{21} readings without any background noise subtraction. This approach ensures that the models can learn to distinguish the tag under both noisy and clean conditions, thereby improving their real-world applicability. The framework then optimises and trains each

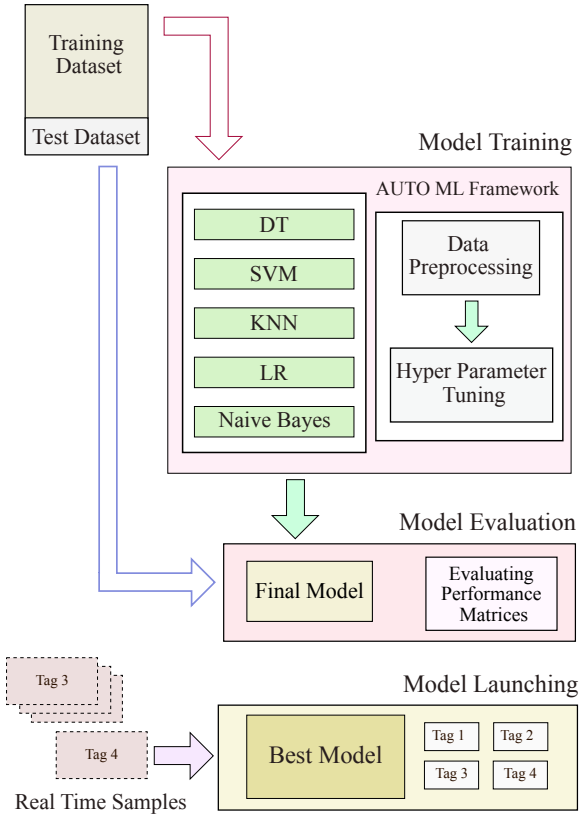

FIGURE 8. The proposed machine learning classification stages, illustrating each phase from data preprocessing to model evaluation and deployment.

TABLE 4. Division of validation data across various combinations of yaw, pitch, roll, and distance parameters.

	Combinations Selected (Ψ, θ, φ, d)
Validation Set 1	Ψ [0°, 40°], θ [0°, 40°], φ [0°, 70°], d [40 cm, 70 cm]
Validation Set 2	Ψ [0°, 40°], θ [0°, 40°], φ [0°, 180°], d [40 cm, 70 cm]
Validation Set 3	Ψ [0°, 40°], θ [0°, 40°], φ [0°, 180°], d [70 cm, 100 cm]
Validation Set 4	Ψ [0°, 40°], θ [0°, 60°], φ [0°, 180°], d [40 cm, 70 cm]
Validation Set 5	Ψ [0°, 60°], θ [0°, 40°], φ [0°, 180°], d [40 cm, 70 cm]
Validation Set 6	Ψ [0°, 70°], θ [0°, 70°], φ [0°, 180°], d [40 cm, 70 cm]

TABLE 5. Accuracy achieved with various validation sets across multiple ML models.

	With background Subtraction					**Without background Subtraction**				
	KNN	**SVM**	**DT**	**LR**	**Naive Bayes**	**KNN**	**SVM**	**DT**	**LR**	**Naive Bayes**
Batch 1	87.72%	95.51%	86.30%	86.10%	44.26%	83.23%	93.47%	87.64%	81.02%	56.39%
Batch 2	91.91%	97.31%	90.47%	89.08%	45.14%	87.20%	94.11%	88.78%	80.07%	55.44%
Batch 3	71.11%	47.31%	30.47%	39.08%	47.14%	67.25%	40.11%	38.12%	47.63%	48.28%
Batch 4	70.11%	78.49%	46.88%	55.45%	50.76%	72.57%	76.80%	47.07%	52.23%	31.19%
Batch 5	47.23%	78.60%	67.56%	65.08%	38.17%	40.71%	75.69%	66.93%	66.76%	32.21%
Batch 6	43.27%	74.27%	56.98%	34.97%	33.74%	44.91%	72.30%	64.85%	55.77%	41.10%

model on these provided datasets, ultimately selecting the best-performing model based on their performance metrics.

To evaluate the trained ML models, the validation dataset which contains variables such as yaw (ψ), pitch (θ), roll (ϕ), and distance (d) is divided into multiple subsets as shown in Table 4. These subsets encompass various combinations of the aforementioned variables, generating multidimensional information. Each of Sets 1 through 6 underwent individual validation using distinct models, and the resulting outcomes are outlined in Table 5. Fig. 9 compares the accuracy achieved by various models on the training dataset for Batch 2, which is identified as the best-classified batch. Upon analyzing the results obtained from various datasets, it is evident that the SVM classifier consistently exhibits satisfactory performance across Set 1 and Set 2.

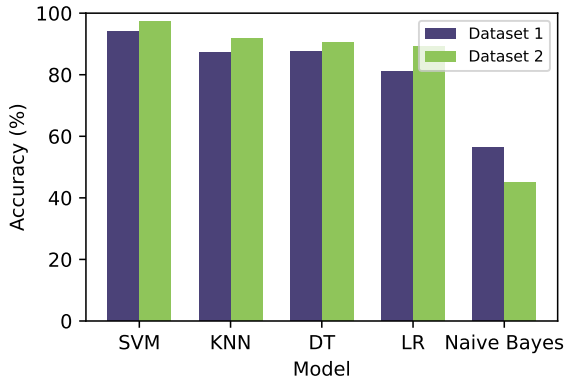

FIGURE 9. Accuracy achieved by various models on the validation dataset Batch 2, identified as the best-classified batch.

TABLE 6. Parameters of the trained SVM model with RBF Kernel.

Parameter	**Optimized Value**
Kernel	RBF
C (Penalty parameter)	10.0
Gamma	0.001

The parameters of the optimized SVM model are shown in Table 6. The SVM model operates by creating hyper-planes that best separate the data points of different classes with the maximum margin. For the RFID dataset, represented as $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i denotes the feature vector corresponding to a tag position, and y_i represents the corresponding labels, the hyperplane can be expressed as:

$$w \cdot x + b = 0 \tag{3}$$

Here, w is the weight vector perpendicular to the hyperplane, x the input feature vector, and b the bias term. The decision boundary is defined by:

$$w \cdot x_i + b \geq 1 \quad \text{if} \quad y_i = 1 \tag{4}$$

$$w \cdot x_i + b \leq -1 \quad \text{if} \quad y_i = 0 \tag{5}$$

These equations describe the margin constraints, where the margin is the distance between the hyperplane and the nearest data point from any class. The goal of the SVM is to maximize this margin while minimizing classification errors.

For handling non-linearly separable data, SVMs use the kernel trick. This technique implicitly maps the input features into a higher-dimensional space where the data can become linearly

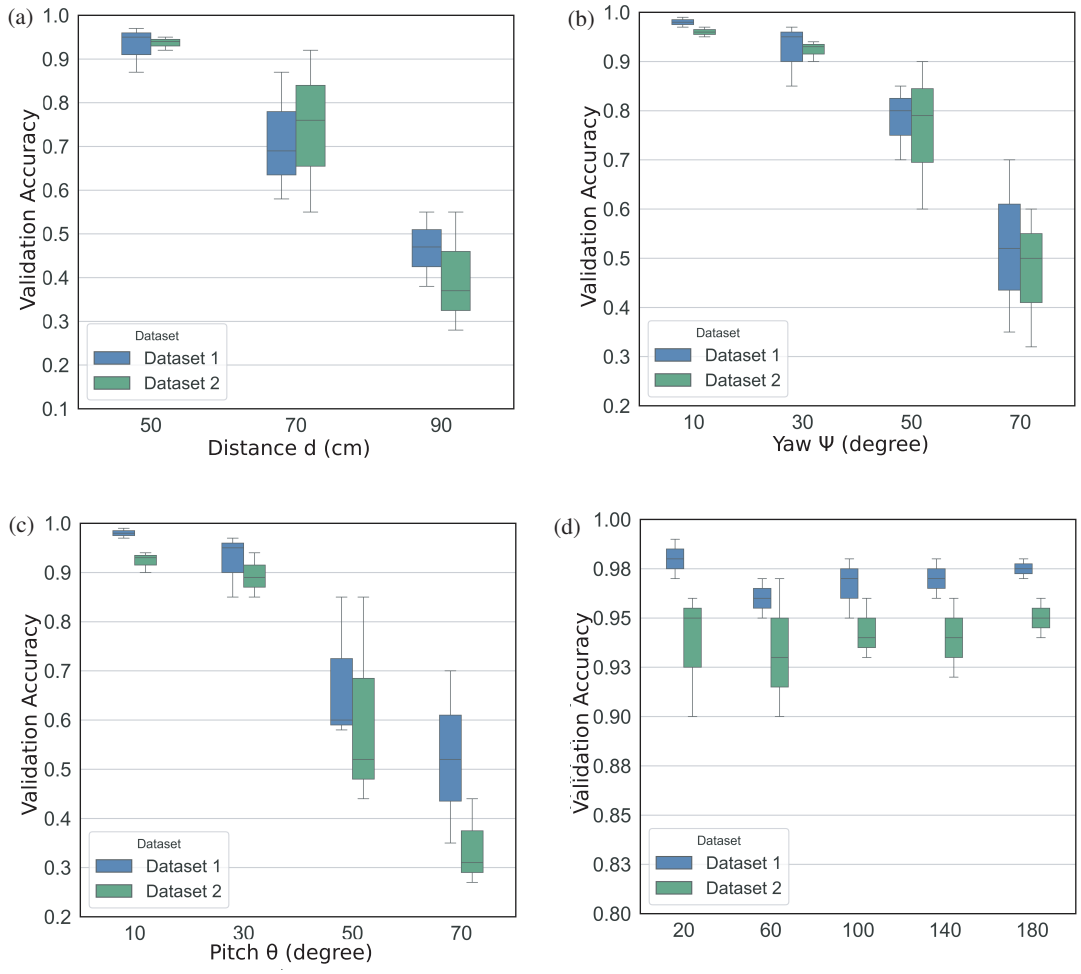

FIGURE 10. Examining the impact of yaw (ψ), pitch (θ), roll (ϕ), and distance (d) parameters on the trained model.

separable. The optimized parameters suggest using the Radial Basis Function (RBF) kernel for higher accuracy with the RFID dataset.

The optimization problem for SVM can be formulated as follows:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{6}$$

subject to: (7)

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n \tag{8}$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, n \tag{9}$$

In this formulation, C is the regularization parameter that balances the trade-off between maximizing the margin and minimizing classification error, while ξ_i are slack variables that allow for some degree of misclassification. By solving this optimization problem, SVM determines the optimal hyperplane that separates the classes in the feature space.

5. MODEL EVALUATION

This section evaluates the performance of the optimised SVM model using various performance matrices. The analysis focuses on examining box plots to gain insights into the model's behaviour across different datasets, particularly under scenarios with and without background noise. Specific focus is placed on the model's classification accuracy about changes in yaw (ψ), pitch (θ), and roll (ϕ) angles, as well as variations in the separation (d) between the interrogator and the tag.

Figure 10(b) presents the performance of the model as the yaw angle is incrementally increased from $0°$ to $70°$. Here, pitch, roll, and distance samples are maintained consistently as outlined in Validation Set 2 (see Table 4). When being tested on Dataset 1, containing background-filtered samples, the model achieved a mean classification accuracy of 97% across the yaw angle range of $0°$ to $40°$. Similarly, the model maintained satisfactory performance on Dataset 2, which comprises raw S_{21} samples, with a mean classification accuracy of 94.1%.

Similarly, in Fig. 10(c), the effect of orientation in pitch angle is assessed keeping the other variables as in validation batch 2. Up to a pitch angle of $40°$, the SVM classifier reliably recognised the tag, irrespective of the presence or absence of

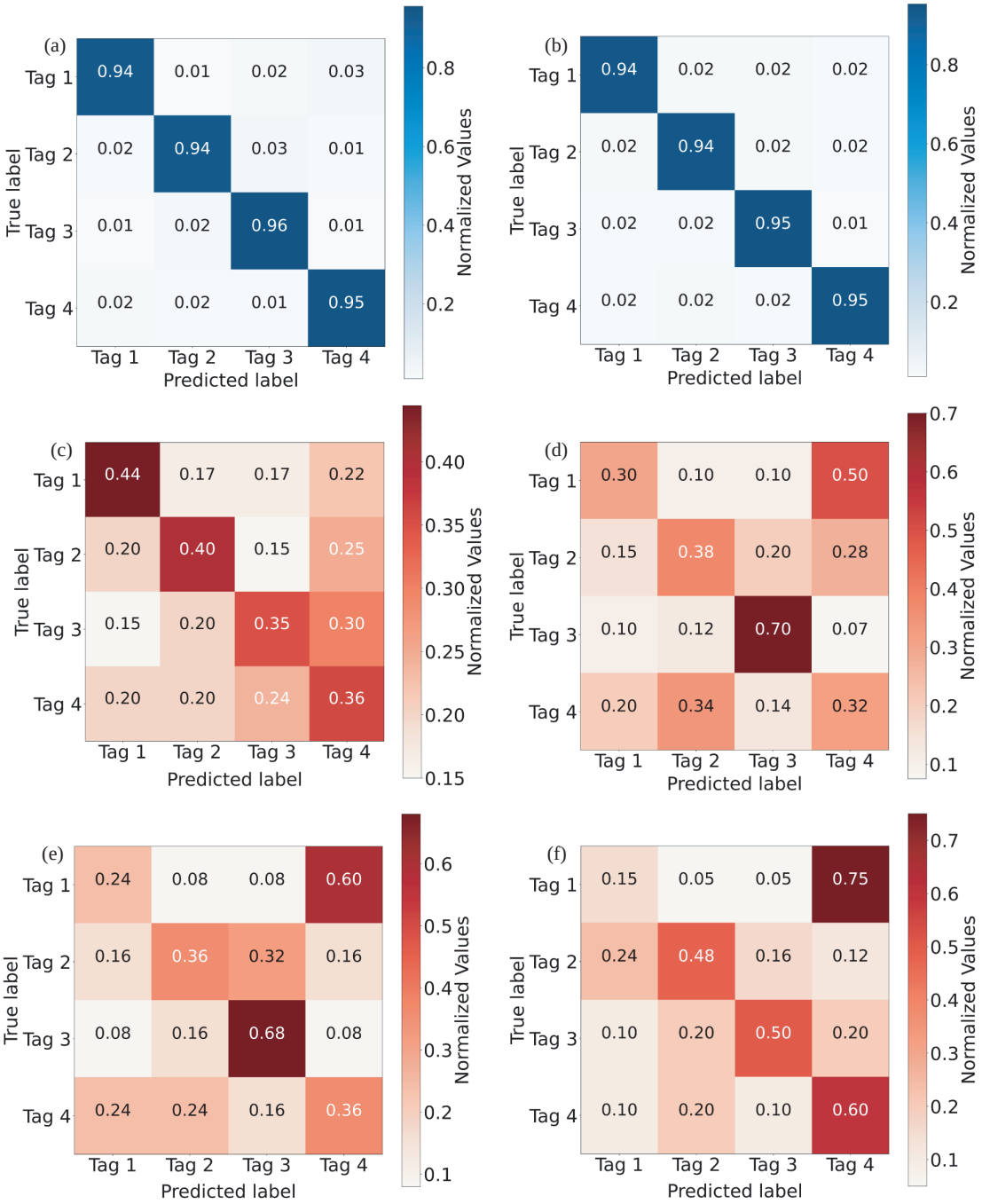

FIGURE 11. Confusion matrix depicting the confidence of the model on each tag for different Validation Sets, utilizing raw S_{21} data without background noise subtraction.

background noise. This indicates the model's efficacy in handling pitch variations within this specific range. However, the model's classification performance began to decline once either the orientation angle yaw or pitch exceeded $40°$. As the angle increases beyond this threshold, the back-scatter signal directed towards the receiving antenna reduces noticeably. This decline in signal strength poses a significant challenge to the model's functionality, ultimately leading to its inability to accurately classify tags. The roll angle has increased from $0°$ to $180°$, as shown in Fig. 10(d). Despite this considerable adjustment in roll angle, the model demonstrates reliable tag identification throughout the roll variations with both datasets. The inherent polarisation independence nature of the tag in the roll plane significantly contributes to this consistent performance, ensuring that the tag's readability is not substantially impacted by changes in roll angle. The distance between the interrogator and the tag is increased while keeping other variables constant, as shown in Fig. 10(a). The model failed to classify the data with acceptable accuracy beyond 70 cm even after subtracting the background noise. This may be because of the small size of

the passive tag, and the backscattered signal from the tag will be limited and not adequate for distinguishing individual tags.

The confusion matrix of various validation batches considered in Section 4 is displayed in Fig. 11. As per Table 5, the SVM model achieved 93.47% validation accuracy on Set 1. Fig. 11(a) shows the corresponding confusion matrix for dataset 1, where each tag is classified to its corresponding classes with minimum error. In contrast, the second validation batch presented different conditions, specifically an increase in the roll angle. Despite this change, Fig. 11(b) shows that the majority of tags are still correctly identified, indicating the model's resilience to variations in the roll angle.

Figure 11(c) confirms that as the distance between the tag and interrogator increases beyond 70 cm, severe misclassification happens among tags. Increasing the interrogator power may help in achieving higher reading ranges. In the experimental sequences labelled Sets 4 through 6, incremental adjustments are made to both yaw and pitch angles. Fig. 11(d) through 11(f) vividly illustrate this, showing that despite being supplied with the maximum available samples, the accuracy of tag classification remained unsatisfactory. These findings are crucial for the development of more robust RFID systems capable of accurate performance in diverse operational environments.

Despite encountering limitations in distinguishing tags at higher orientation angles or distances, the model exhibited consistent and reliable classification within specific ranges. Notably, the model demonstrated proficiency in accurately classifying tags positioned at distances ranging from 40 cm to 70 cm. Moreover, it effectively discerned tags across a yaw range spanning from $0°$ to $40°$, Pitch from $0°$ to $40°$, and Roll from $0°$ to $180°$ showcasing its robustness in classifying tags across a considerable angular span.

The precision, recall, and F1 score matrices of the model on Set 2, regarded as the best-classified set, are depicted in the accompanying Fig. 12. These metrics collectively signify the model's high levels of accuracy, thoroughness, and sensitivity in accurately categorising tags within the dataset. Such consistent and robust performance across multiple evaluation criteria underscores the effectiveness and reliability of the classification system employed, reflecting positively on its overall performance and suitability for the intended application.

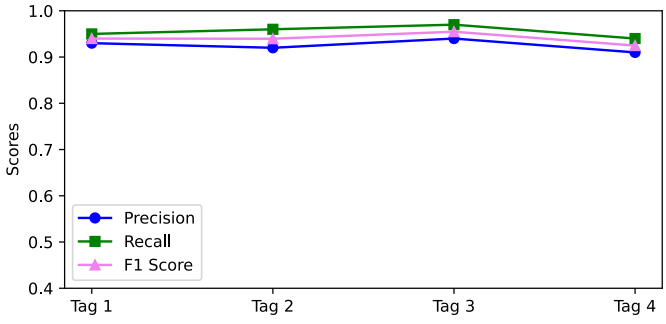

FIGURE 12. Estimation of precision, recall, and F1 score considering validation set 1 and raw S_{21} data.

6. CONCLUSION

This study explores the feasibility of employing ML techniques to improve accuracy and reliability of CRFID systems in real-world environments. The primary challenge encountered is the limited availability of comprehensive RFID datasets. To address this, an automated tag positioning system with an SCPI interface is implemented to capture backscattered signals from various tag orientations. A diverse dataset is generated by interrogating the CRFID tag while varying several parameters: 1) distance (d) from 40 cm to 100 cm, 2) roll (ϕ) from $0°$ to $180°$, 3) pitch (θ) from $0°$ to $70°$, and 4) yaw (ψ) from $0°$ to $70°$. The generated datasets are trained and validated using SVM, k-NN, DT, LR, and Naive Bayes approaches. Among these models, SVM model demonstrated superior overall performance. Specifically, the model successfully extracted tag information up to 70 cm from the interrogator, with yaw and pitch orientations ranging from $0°$ to $40°$ and roll orientations from $0°$ to $180°$. The methodology described in this study is both scalable and generalisable, capable of adapting to any number of tags or bits within varying environments or noise conditions.

ACKNOWLEDGEMENT

The authors would like to express their deep gratitude to the University Grants Commission, India, for providing essential funding for this research.

REFERENCES

[1] Mulloni, V. and M. Donelli, "Chipless RFID sensors for the Internet of Things: Challenges and opportunities," *Sensors*, Vol. 20, No. 7, 2135, 2020.

[2] Hassan, W., T. M. Ali, and A. Attiya, "Near field RFID tag for IoT in sub-six GHz band," *Progress In Electromagnetics Research Letters*, Vol. 100, 169–175, 2021.

[3] Subrahmannian, A. and S. K. Behera, "Chipless RFID: A unique technology for mankind," *IEEE Journal of Radio Frequency Identification*, Vol. 6, 151–163, 2022.

[4] Khan, M. U. A., R. Raad, J. Foroughi, M. S. Raheel, and S. Houshyar, "An octagonal-shaped conductive HC12 & LIBERATOR-40 thread embroidered chipless RFID for general IoT applications," *Sensors and Actuators A: Physical*, Vol. 318, 112485, 2021.

[5] Ahmadihaji, A., R. Izquierdo, and A. Shih, "From chip-based to chipless rfid sensors: A review," *IEEE Sensors Journal*, Vol. 23, No. 11, 11 356–11 373, 2023.

[6] Babaeian, F. and N. C. Karmakar, "Time and frequency domains analysis of chipless RFID back-scattered tag reflection," *IoT*, Vol. 1, No. 1, 109–127, 2020.

[7] Brinker, K. R. and R. Zoughi, "A review of chipless RFID measurement methods, response detection approaches, and decoding techniques," *IEEE Open Journal of Instrumentation and Measurement*, Vol. 1, 1–31, 2022.

[8] Nijas, C. M., U. Deepak, P. V. Vinesh, R. Sujith, S. Mridula, K. Vasudevan, and P. Mohanan, "Low-cost multiple-bit encoded chipless RFID tag using stepped impedance resonator," *IEEE Transactions on Antennas and Propagation*, Vol. 62, No. 9, 4762–4770, 2014.

[9] Jalil, M. E. B., M. K. A. Rahim, H. Mohamed, N. A. B. Samsuri, N. A. Murad, R. Dewan, H. B. A. Majid, N. B. M. Nafis, L. O. Nur, and B. S. Nugroho, "High capacity and miniaturized flexi-

ble chipless RFID tag using modified complementary split ring resonator," *IEEE Access*, Vol. 9, 33 929–33 943, 2021.

[10] Herrojo, C., F. Paredes, J. Mata-Contreras, and F. Martín, "Chipless-RFID: A review and recent developments," *Sensors*, Vol. 19, No. 15, 3385, 2019.

[11] Braga-Neto, U., *Fundamentals of Pattern Recognition and Machine Learning*, Springer, 2020.

[12] Amiri, Z., A. Heidari, N. J. Navimipour, M. Unal, and A. Mousavi, "Adventures in data analysis: A systematic review of deep learning techniques for pattern recognition in cyber-physical-social systems," *Multimedia Tools and Applications*, Vol. 83, No. 8, 22 909–22 973, 2024.

[13] Serey, J., M. Alfaro, G. Fuertes, M. Vargas, C. Duran, R. Ternero, R. Rivera, and J. Sabattin, "Pattern recognition and deep learning technologies, enablers of industry 4.0, and their role in engineering research," *Symmetry*, Vol. 15, No. 2, 535, 2023.

[14] Farahani, M. A., M. R. McCormick, R. Gianinny, F. Hudacheck, R. Harik, Z. Liu, and T. Wuest, "Time-series pattern recognition in Smart Manufacturing Systems: A literature review and ontology," *Journal of Manufacturing Systems*, Vol. 69, 208–241, 2023.

[15] Su, C., X. Wang, C. Zou, L. Jiao, and Y. Tao, "An innovative method based on wavelet analysis for chipless RFID tag detection," *Electronics*, Vol. 13, No. 12, 2375, 2024.

[16] Aliasgari, J. and N. C. Karmakar, "Mathematical model of chipless RFID tags for detection improvement," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 68, No. 10, 4103–4115, 2020.

[17] Chen, Y., F. Zheng, T. Kaiser, and A. J. H. Vinck, "An information-theoretic approach to the chipless RFID tag identification," *IEEE Access*, Vol. 7, 96 984–97 000, 2019.

[18] Wei, F., W. T. Li, X. W. Shi, and Q. L. Huang, "Compact UWB bandpass filter with triple-notched bands using triple-mode stepped impedance resonator," *IEEE Microwave and Wireless Components Letters*, Vol. 22, No. 10, 512–514, 2012.

[19] Wei, F., Q. Y. Wu, X. W. Shi, and L. Chen, "Compact UWB bandpass filter with dual notched bands based on SCRLH resonator," *IEEE Microwave and Wireless Components Letters*, Vol. 21, No. 1, 28–30, 2011.

[20] Jeong, S., J. G. D. Hester, W. Su, and M. M. Tentzeris, "Read/interrogation enhancement of chipless RFIDs using machine learning techniques," *IEEE Antennas and Wireless Propagation Letters*, Vol. 18, No. 11, 2272–2276, 2019.

[21] Arjomandi, L. M. and N. C. Karmakar, "An enhanced chipless RFID system in 60 GHz using pattern recognition techniques," in *2018 48th European Microwave Conference (EuMC)*, 973–976, Madrid, Spain, Sep. 2018.

[22] Nastasiu, D., R. Scripcaru, A. Digulescu, C. Ioana, R. de Amorim, Jr., N. Barbot, R. Siragusa, E. Perret, and F. Popescu, "A new method of secure authentication based on electromagnetic signatures of chipless RFID tags and machine learning approaches," *Sensors*, Vol. 20, No. 21, 6385, 2020.

[23] Sokoudjou, J. J. F., F. Villa-González, P. García-Cardarelli, J. Díaz, D. Valderas, and I. Ochoa, "Chipless RFID tag implementation and machine-learning workflow for robust identification," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 71, No. 12, 5147–5159, 2023.

[24] Betancourt, D., K. Haase, A. Hübler, and F. Ellinger, "Bending and folding effect study of flexible fully printed and late-stage codified octagonal chipless RFID tags," *IEEE Transactions on Antennas and Propagation*, Vol. 64, No. 7, 2815–2823, 2016.

[25] Rohde & Schwarz, "Rohde & schwarz zvb 20 vector network analyzer," [Online]. Available: https://www.testequipmenthq.com/datasheets/Rohde-Schwarz-ZVB20-Datasheet.pdf, 2020.

[26] Sharma, S. and V. Sharma, "Performance of various machine learning classifiers on small datasets with varying dimensionalities: A study," *Circulation in Computer Science*, Vol. 1, No. 1, 30–35, 2016.

[27] Brigato, L. and L. Iocchi, "A close look at deep learning with small data," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2490–2497, Milan, Italy, Jan. 2021.

[28] Yelure, B. S., S. V. Patil, S. B. Patil, and S. B. Nemade, "Solving multi-class classification problem using support vector machine," in *2022 International Conference on Futuristic Technologies (INCOFT)*, 1–4, Belgaum, India, Nov. 2022.

[29] Ghosh, D. and J. Cabrera, "Enriched random forest for high dimensional genomic data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 19, No. 5, 2817–2828, 2021.

[30] Mladenova, T. and I. Valova, "Comparative analysis between the traditional K-Nearest Neighbor and Modifications with Weight-Calculation," in *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 961–965, Ankara, Turkey, Oct. 2022.

[31] Salehin, I., M. S. Islam, P. Saha, S. M. Noman, A. Tuni, M. M. Hasan, and M. A. Baten, "Automl: A systematic review on automated machine learning with neural architecture search," *Journal of Information and Intelligence*, Vol. 2, No. 1, 52–81, 2024.