

Review

Comprehensive Survey on Radar Systems and Its Target Classification Techniques

Paramveer Singh^{1*}, Shree Menkudale^{1*} , Samraddhi Soni², Vanita Raj Tank¹ , A. A. Brazil Raj²

¹Department of Electrical and Electronics Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, India

²RF/Photonics Laboratory, Departments of Electronics Engineering, Defence Institute of Advance Technology, Pune, India

E-mail: paramveers9451@gmail.com (P.S.); shreedmenkudale@gmail.com (S.M.)

Received: 16 October 2024; **Revised:** 16 January 2025; **Accepted:** 17 January 2025

Abstract: This paper surveys target classification techniques in radar systems, focusing on the transformative role of artificial intelligence in enhancing detection and classification capabilities. It reviews the evolution of radar architectures, emphasizing their design, functionality, and key parameters that drive performance. The study spans a range of approaches, from traditional machine learning to advanced deep learning methods, including CNNs, RNNs, self-attention mechanisms, vision transformers, and 2D-SPS. These innovations enable breakthroughs in micromotion detection, background noise reduction, and prediction accuracy. By highlighting applications across various industries, this work provides valuable insights to researchers and engineers, paving the way for advancements in radar technology driven by robust hardware and sophisticated algorithms.

Keywords: CNN, RNN, vision transformer, MHSA, micro-doppler signatures, 2D-power spectrum, residual learning, radar

Acronyms

| | | | |
|-------|-------------------------------------|------|---|
| LiDAR | Light Detection and Ranging | UAV | Unmanned Aerial Vehicle |
| RADAR | Radio Detection and Ranging | ViT | Vision Transformer |
| LFM | Linear Frequency Modulation | STFT | Short-Time Fourier Transform |
| RNN | Recurrent Neural Network | PWVD | Pseudo-Wigner-Ville Distribution |
| MHSA | Multi-Head Self-Attention | SPS | Singularity Power Spectrum |
| SAR | Synthetic Aperture Radar | GAP | Global Average Pooling |
| FMCW | Frequency-Modulated Continuous Wave | ROC | Receiver Operating Characteristic |
| DDS | Direct Digital Synthesis | SINR | Signal-to-Interference-plus-Noise Ratio |
| CNN | Convolutional Neural Network | DP | Dynamic Programming |
| SF-CW | Step-Frequency Continuous Wave | EM | Electromagnetic |
| PRI | Pulse Repetition Interval | MIMO | Multiple-Input Multiple-Output |
| PRF | Pulse Repetition Frequency | DOA | Direction of Arrival |
| TFD | Time Frequency Distribution | NLP | Natural Language Processing |
| RTI | Range-Time Intensity | SFW | Stepped Frequency Waveform |

Copyright ©2025 Paramveer Singh, et al.

DOI: <https://doi.org/10.37256/jeee.4120255934>

This is an open-access article distributed under a CC BY license

(Creative Commons Attribution 4.0 International License)

<https://creativecommons.org/licenses/by/4.0/>

| | | | |
|-------|---|-------|--|
| IDFT | Inverse Discrete Fourier Transform | OFDM | Orthogonal Frequency Division Multiplexing |
| FFT | Fast Fourier Transform | GRU | Gated Recurrent Unit |
| MFCC | Mel-Frequency Cepstral Coefficients | CPI | Coherent Processing Interval |
| PCA | Principal component analysis | DFT | Discrete Fourier Transform |
| FFN | Feed Forward Neural Network | GELU | Gaussian Error Linear Unit |
| LN | Layer Normalization | CLS | Classification |
| MSFCA | Multi-Scale Filtering and Channel Attention | FOPEN | Foliage Penetration |

1. Introduction

Radar systems have become essential in applications such as automotive safety, healthcare, surveillance, and environmental monitoring. Their ability to function reliably under diverse weather and lighting conditions, unlike technologies such as cameras and lidar, makes them highly versatile. By using radio waves, radars overcome limitations posed by environmental factors like fog or darkness, offering a robust solution for real-time sensing. Recent advancements in radar technology have also enabled compact, cost-effective, and high-resolution systems, further enhancing their utility. A key challenge for radar systems is accurate target classification, which is vital for applications like collision avoidance, activity monitoring, and defence. This challenge stems from the complexity of signal reflections, interference, and clutter in real-world scenarios. Various methods have been developed to address this, including traditional signal processing for feature extraction, machine learning for predictive modeling, and deep learning for direct feature extraction and classification. While deep learning has shown superior performance, it often demands high computational resources and large datasets. This review transitions to the background and history of radar technology, tracing its evolution from early 20th-century concepts to today's advanced systems. Initially developed for military use during World War II, radar has since incorporated digital signal processing and phased-array technology, continuously adapting to meet the needs of various applications. This historical progression highlights the key advancements that have enabled the development of modern radar-based target classification techniques.

2. Background and history

Radio Detection and Ranging, abbreviated as RADAR, plays an important role in modern defence systems, air traffic control, weather forecasting, and autonomous vehicle technology, providing critical capabilities in detecting and tracking objects over long distances in various environmental conditions. They are much more advanced nowadays and are also used for target imaging and classification. Radar imaging involves range and cross-range profiling, as shown in Figure 1, to capture target details, with resolution affected by pulse width. Shorter pulses or modulation techniques enhance accuracy, while Synthetic Aperture Radar (SAR) uses Doppler shifts for improved cross-range resolution, enabling high-resolution imaging for applications like terrain mapping and surveillance. Antenna pattern distortion complicates this, as shown by Gaurav et al. [1, 2], while Kumawat et al. [3] proposed methods to reduce interference and improve multi-target image accuracy. Advanced techniques like pulse compression and polarization scattering refine imagery, aiding automatic target recognition. All radar systems, from weather to military, share core functions—searching, tracking, and imaging. Radar creates 2D images of 3D objects by collecting reflection data from various angles, which is then mapped to a 2D plane, with brighter points indicating higher reflectivity. The goal is to train robust CNN models for real-world applications. This expansion in capability extends their applications to subsurface mapping, healthcare, identifying buried or hidden objects, and accurately recognizing them. The ability to effectively ascertain target attributes allows for timely and informed decision-making, thereby enhancing operational efficiency and safety in multiple domains. Consequently, target classification and the determination of attributes such as range, velocity, and angle of approach continue to be dynamic and fast-evolving fields of research, with ongoing developments and expanding applicability across various tasks.

Table 1 provides a structured overview of the key milestones and concepts in the development of radar technology. It highlights the contributions of various researchers and institutions, as well as the evolution of radar from its theoretical foundations to modern advanced systems.

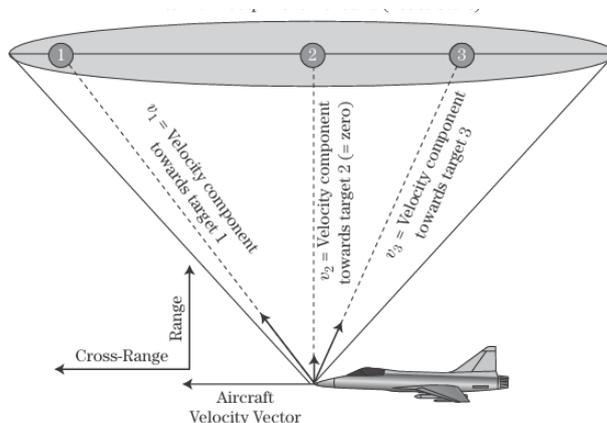


Figure 1. Imaging of an aircraft using radars

Table 1. Developments in radar technology

| Year | Contributor(s) | Key contribution/concept | Significance/Impact |
|-------|---|---|--|
| 1865 | James C. Maxwell [4] | Work on EM wave propagation | Laid the theoretical foundation for radar. |
| 1886 | Heinrich Hertz [5] | Experimentally confirmed the existence of EM waves | Provided the first step toward applying EM principles in radar. |
| 1904 | Christian hulsmeyer [6] | Detected a ship at a range of one mile | First practical demonstration of radar principles. |
| 1915 | Robert W. Watt [7] | Used radio signals to detect thunderstorms | Early application of radio waves for detection. |
| 1920s | Marconi [8] | Pioneering work in radio communications | Influenced early radar development. |
| 1925 | Tuve et al. [9] | Experiment contributing to understanding radar's range-measuring capabilities | Furthered understanding of radar functionality. |
| 1928 | Henry Nyquist, Claude E. Shannon | Nyquist-Shannon sampling theorem | Prevents signal distortion during sampling |
| 1930 | Glaser et al. [10] | Demonstrated Hyland's work on aircraft detection capabilities. | Highlighted radar's significance in defense and aviation. |
| 1940 | MIT radar laboratory [11] | Development of the magnetron (1 kW, 3 GHz) | Significantly enhanced radar power and capabilities. |
| 1940s | Stanford university | Invention of the klystron amplifier | Provided a powerful tool for amplifying high-frequency signal |
| 1943 | Van vleck | Matched filtering | Boosts signal detection, reduces noise and improves resolution |
| 1944 | Luiz alvarez | Phased-array radar | Enabled electronic beam steering, enhanced radar accuracy |
| 1945 | Herald cramer and Calvin R. Rao | Cramer-Rao lower bound | Sets limits on parameter (range, velocity etc.) estimation accuracy in radar systems |
| 1948 | Shannon [12] | Information theory and channel capacity ($C = \max I(X; Y) p(X)$) | Provided theoretical framework for assessing radar performance and underpins signal processing. |
| 1950s | Klauder, J. R., Price, A. C., Darlington, S., & Albersheim, W. J. | Pulse compression | Achieves high range resolution with long pulses, increasing average transmitted power. |
| 1954 | Swerling [13] | Target models (Swerling I-IV) explaining RCS variations | Improved radar detection capabilities by accounting for target complexity. |
| 1960s | Cooley and Tukey [14] | Fast fourier transform (FFT) algorithm | Revolutionized signal processing by enabling rapid computation of the discrete fourier transform. |
| 1967 | Brown and William M. | Synthetic Aperture Radar (SAR) | Uses the motion of the radar platform to synthesize a large antenna aperture, achieving high along-track resolution. |
| 1968 | Ruttenburg et al. [15] | Stepped frequency waveform (SFW) | Improved range resolution through inter-pulse phase coding. |
| 1979 | Schaefer, R., Schafer, R., & Mersereau, R. | Doppler processing | Measures the change in frequency of the returned signal to determine target velocity. |
| 1986 | Schmidth et al. [16] | MUSIC (Multiple Signal Classification) algorithm | High-resolution direction of arrival (DOA) estimation. |

Table 1. Cont.

| | | | |
|-------|--------------------------|---|---|
| 1989 | Roy et al. [17] | ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) algorithm | Computationally efficient DOA estimation. |
| 2003 | Bliss et al. [18] | MIMO radars | Improved spatial resolution and interference resilience. |
| 2006 | S. Haykin | Cognitive radar | Pioneered new radar paradigm |
| 2010 | A. Hassanien et al. [19] | Phased-MIMO radar | Combined advantages of phased array and MIMO systems. |
| 2020 | Fortunati et al. [20] | Use of large number of antennas | Improved detection accuracy and spatial resolution. |
| 2020 | Latpate et al. [21] | C-band continuous wave (CW) radar system with micro-Doppler extraction | Enhanced detection of low RCS targets and ability to distinguish approaching from receding targets. |
| 2020s | Farlik et al. [22] | Surveys of radar techniques for UAV detection and imaging | Highlighted the importance of Doppler effect, advancements for LSS targets, and the use of deep learning. |
| 2023 | Sampurna et al. [23] | Photonics-based radars | Low cost and power consumption, lightweight and reduced EM interference. |
| 2023 | Lakshmi et al. [24] | Fuzzy logic for channel response prediction in optical networks | Accurate (92.1%) prediction of environmental effects on channel response. |

3. Advanced radar designs and mathematical models

There are various types of radar, classified based on waveforms, antenna configurations, and signal processing techniques. The subsequent hierarchy is shown in Figure 2. Understanding multiple radar designs is crucial for accurate target classification, as each type excels in specific scenarios. For instance, waveform diversity enhances detection precision, monostatic and bistatic configurations affect spatial coverage, and advanced signal processing techniques improve resolution and interference mitigation. By leveraging these strengths, researchers optimize performance for complex and dynamic environments, ensuring reliable and precise classification across diverse applications.

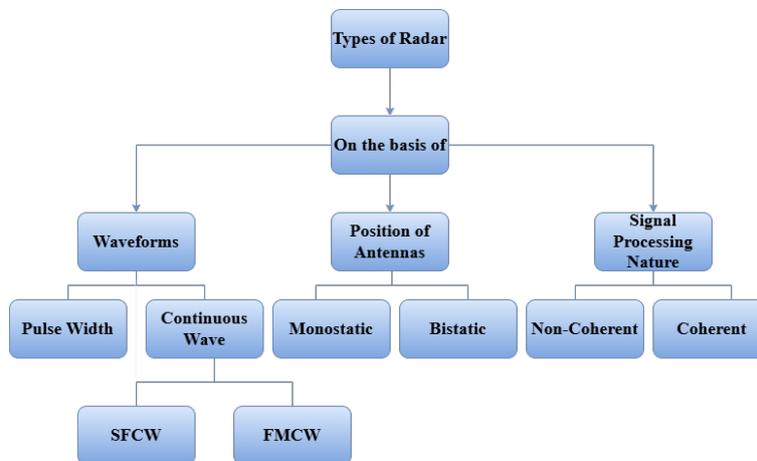


Figure 2. Types of radar

3.1 FM-CW radar

Kanzarkar et al. [25] study the design and digital architecture of frequency-modulated continuous wave (FMCW) radar, whose transmitted frequency is linearly dependent on time, which is enhanced by direct digital synthesis (DDS) technology. The study demonstrates how DDS improves the precision of sensing and range-detecting capabilities by allowing for variable and accurate frequency modulation. DDS enables real-time waveform synthesis, which is critical for changing radar signal characteristics in dynamic settings, resulting in higher precision in range and velocity measurements. This integration enables real-time photography, which captures high-resolution data necessary for applications that require quick

response, hence boosting safety through exact measurements. FMCW radar's architecture consists of signal generators, antennas, receivers, and processors, with DDS playing a critical role in signal production, resulting in better image resolution and system adaptability across diverse operational conditions. Beyond typical automotive and aerospace uses, FMCW radar is being tested in healthcare [26] for patient mobility tracking and non-invasive sensing, demonstrating its versatility and value in a variety of disciplines as discussed by Singh et al. [27] who proposed photoplethysmography technique. However, issues exist regarding the cost-effectiveness of DDS technology, which must be solved to promote greater use and accessibility across a wide range of businesses. Combining FMCW radar with other sensing technologies, such as LiDAR and cameras, improves multi-modal sensing capabilities, particularly in autonomous cars. FMCW radars continuously transmit, and receive signals, unlike pulsed radar, from the target. This echo signal, received at the receiver antenna, is analyzed for detection of objects and their relative motion by exploiting the change due to the Doppler effect. The echo signal is time-shifted transmitted signal when scattered from a stationary target. But when it strikes a moving target, its frequency either increases or decreases, to the relative motion of the target towards or away from the radar. This change in frequency is shown in Figure 3 and is termed as Doppler Shift [28]. This shift is exploited to find the relative radial velocity of the target. This modulation correlates with the time delay of the echo signal, thereby, allowing precise range measurement; useful in automotive radar systems, for collision avoidance, where range resolution is critical.

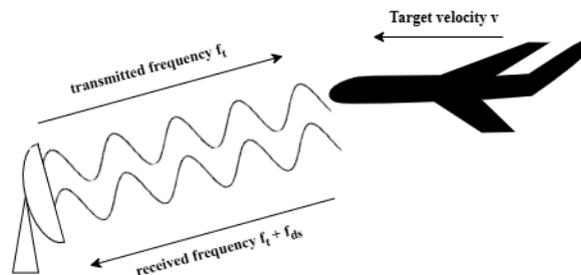


Figure 3. Detection of moving objects using CW radar

Studying autodyne, homodyne, and heterodyne radar configurations of FMCW radar is essential because each offers unique advantages in processing and interpreting radar signals for specific applications. These configurations determine how the transmitted and received signals are mixed and processed, directly influencing the resolution, sensitivity, and noise performance of the system.

Autodyne: This configuration, shown in Figure 4, is one of the simplest configurations. In this setup, the same oscillator is used for both the transmission and reception of the signal. The transmitted and received signals are directly mixed, allowing the radar to detect the Doppler shift without the need for an external mixer. This makes the autodyne radar [29] ideal for low-cost, low-power applications. However, the drawback is that it often suffers from higher noise levels and limited sensitivity, making it less suitable for applications requiring high precision.

Homodyne: The homodyne configuration improves upon the autodyne design by using a separate local oscillator for the mixing process. In this setup, as Figure 5 shows, the received and local oscillator signals (typically of the same frequency as transmitted signal) are mixed. This approach allows for better isolation between the transmit and receive paths, resulting in lower noise levels and improved sensitivity. Homodyne radars are commonly used in applications where the balance between complexity and performance is critical, offering a good compromise between cost and effectiveness.

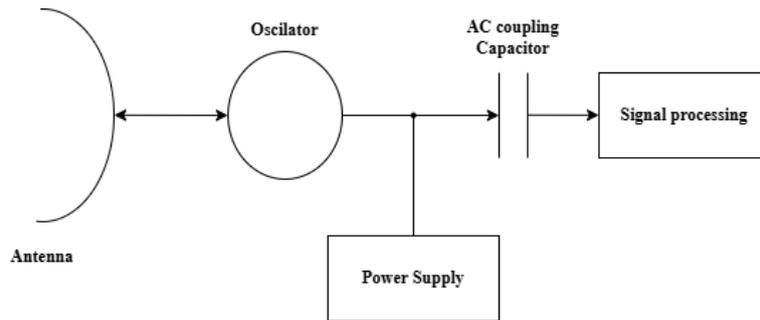


Figure 4. Block diagram for autodyne configuration

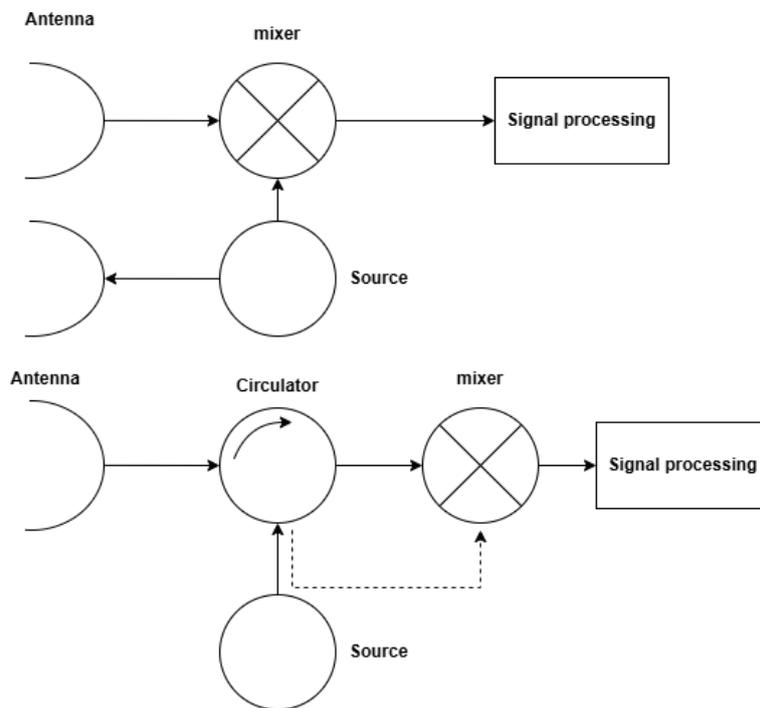


Figure 5. Block diagram for homodyne configuration

Heterodyne: This configuration represents a more advanced and complex approach to radar design. In this system, the transmit modulation is applied at an intermediate frequency (IF) rather than directly at the carrier frequency. Then, the modulated IF and local oscillator signal is mixed to produce the desired carrier frequency for transmission. On the receiving side, the echo and local oscillator signal is again mixed, as shown in Figure 6, and the signal is processed at the intermediate frequency. The primary advantage of the heterodyne configuration is its ability to separate the received signal frequency from the transmit frequency. This separation reduces the impact of amplitude and phase modulation noise from the transmitter, resulting in improved performance, particularly in terms of sensitivity and signal-to-noise ratio.

However, this configuration also introduces greater complexity, making it more challenging to implement and more costly than the simpler autodyne and homodyne systems. One of the key benefits of heterodyne configuration is the improved signal processing capabilities that come with operating at an intermediate frequency. By shifting the signal processing away from the carrier frequency, the radar system can reduce the noise power associated with the transmitted signal, particularly in terms of amplitude and phase noise. This noise reduction enhances the radar's ability to detect weak signals and improves overall system performance. FM-CW radar systems are particularly advantageous in applications requiring compact, short-range, and cost-effective solutions. These systems are often implemented using solid-state

technology, making them reliable and efficient. Their widespread use spans various fields, from industrial applications such as liquid level measurement in storage tanks, illuminators in fire control systems, short-range navigation, battlefield surveillance, altimeters, proximity fuses, and even automobile cruise control systems. As the automotive industry moves towards making radar systems standard in new vehicles, this technology is poised to become one of the most ubiquitous radar variants in everyday life.

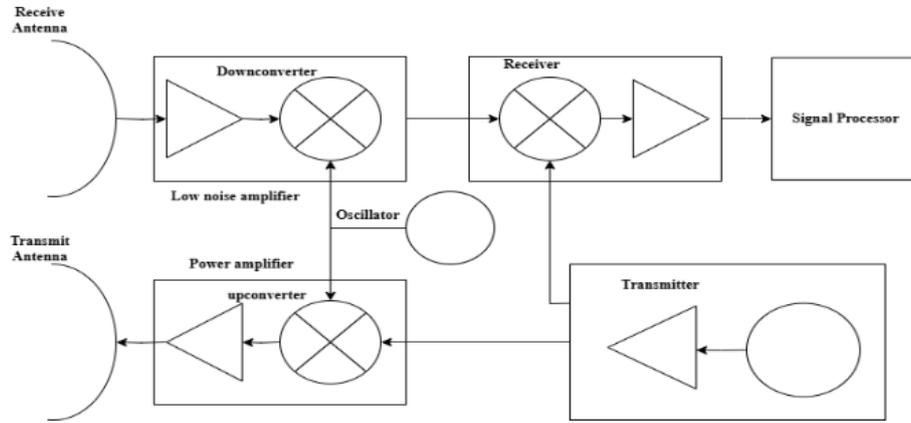


Figure 6. Block diagram for heterodyne configuration

3.2 SF-CW radar

Step-frequency CW radar is a type of radar that transmits a sequence of continuous waves with varying frequencies. It steps through a range of frequencies in discrete increments, which allows the radar to cover a broad bandwidth. This technique enables high-resolution range measurements and precise target detection by analyzing the reflected signals at different frequencies. Using stepped-frequency chirps to improve resolution, Arya et al. presents two novel approaches to data processing for reducing radio frequency interference (RFI) in Ultra-Wide Band radar systems. A band-rejection level of 40–50 dB is achieved and target detection is maintained with minimal side lobe distortion (maximum 2 dB at 12.5% band nulling) using the first method, matched filtering with virtual filtering, which uses multiple narrowband chirps centered at different frequencies.

An RFI detector detects and nullifies chirp data affected by significant interference. To achieve a band-rejection level of 30–40 dB, the second method—spread processing with virtual filtering—involves de-chirping received segments, nulling those that surpass a threshold RFI level, and integrating the remaining segments. Nevertheless, this approach suffers from more side lobe distortion (5–10 dB at 13.33% band nulling), but it saves money by lowering the target return amplitude. Both approaches outperform conventional hardware-based RFI suppression techniques in some important ways, including easier implementation via data processing and improved system performance. An SF-CW radar transmits a signal composed of K sinusoidal frequencies, ranging from f_0 to $f_0 + (K - 1)\Delta f$, where Δf is the frequency step size as depicted in Figure 7. The bandwidth B of the radar is given by $B = K \cdot \Delta f$. Each frequency tone is transmitted for a time interval T , and the total transmit time of a single waveform is $T_K = K \cdot T$.

A single transmit waveform at a time t is expressed as:

$$x(t) = \sum_{k=0}^{K-1} \text{rect}\left(\frac{t - \frac{T}{2} - kT}{T}\right) e^{-j2\pi(f_0 + k\Delta f)t} \quad (1)$$

The received signal at the slow time index l and receiver antenna m is modeled by:

$$s_l(t, m) = \sum_{p=1}^P \gamma_{pm} x(t - \tau_{pm}(l)) + \varepsilon_m \quad (2)$$

where γ_{pm} denotes the amplitude of the p -th target, ε_m models clutter, and $\tau_{pm}(l)$ is the time delay of the received signal, which is decomposed into:

$$\tau_{pm}(l) = \tau_d(pm) + \tau_v p(l) \quad (3)$$

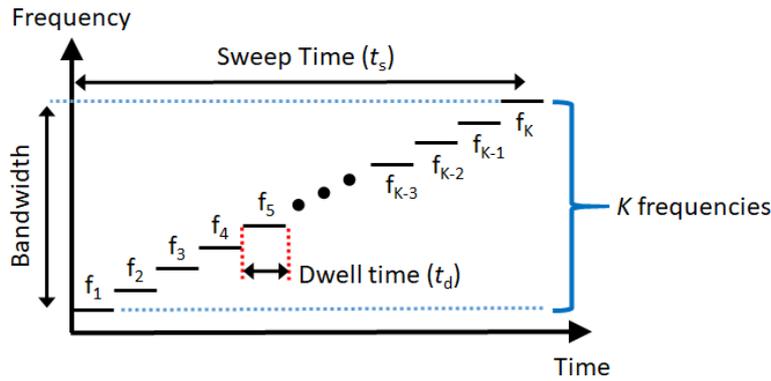


Figure 7. Depiction of the single chirp of SF-CW radar

The dominant component $\tau_d(pm)$ relates to the distance between the radar and the target, given by:

$$\tau_d(pm) = \frac{1}{c} (2d_p + m\delta \sin(\theta_p)) \quad (4)$$

where c is the speed of light, δ is the inter-antenna distance, d_p is the range, and θ_p is the angle. After discretization and conversion to baseband, the normalized signal is:

$$\tilde{s}_l(k, m) = \sum_{p=1}^P e^{-j2\pi(f_0 + k\Delta f)\tau_{pm}(l)} + \tilde{\varepsilon}_m \quad (5)$$

The range profile is computed using an N -point inverse fourier transform over the fast time-frequency index k :

$$s_l(n, m) = \frac{1}{K} \sum_{k=0}^{K-1} \tilde{s}_l(k, m) e^{j2\pi \frac{kn}{N}} \quad (6)$$

which simplifies to:

$$s_l(n, m) = \frac{1}{K} \sum_{p=1}^P e^{-j2\pi f_0 \tau_{pm}(l)} \frac{\sin\left(\frac{\pi K}{N} (n - N\Delta f \tau_{pm}(l))\right)}{\sin\left(\frac{\pi}{N} (n - N\Delta f \tau_{pm}(l))\right)} e^{j\pi \frac{K-1}{N} (n - N\Delta f \tau_{pm}(l))} \quad (7)$$

The range profile is then: $r_l(n, m) = |s_l(n, m)|$ with a maximum unambiguous range d_{\max} , range resolution Δd , and range measurement granularity δd given by

$$d_{\max} = \frac{c}{2\Delta f} \quad (8)$$

$$\Delta d = \frac{c}{2B} \quad (9)$$

$$\text{and } \delta d = \frac{c}{2\Delta f N} \quad (10)$$

where Δd denotes the minimum distance between distinguishable targets. By choosing $N > K$, the granularity δd increases, improving the sampling resolution of the range profile. Stepped linear frequency modulation (LFM) technique is an innovative way to achieve high-range resolution in phased array radar systems, as described by Arya et al. [30]. He suggests two techniques to create a wide-band equivalent response from narrow-band chirp segments, stretch processing and matched filtering. Narrow-band alternatives are required because traditional wide-band signals cause problems such as beam squinting in active phased array radars and grating lobe formation.

3.3 Pulsed radar

Pulsed radars send EM waves out for a very short time, between 0.1 to 10 μ s, but sometimes it is more or less depending upon the application. During this pulse, the receiver cannot detect any pulse as it is disconnected from the antenna to protect it from the high-power waves. After the pulse is sent, the receiver reconnects to the antenna to listen for any echoes from objects. This listening time, combined with the pulse duration, makes up one full radar cycle, known as the pulse repetition interval abbreviated as PRI.

The pulse repetition frequency is how many transmit/receive cycles the radar completes per second, measured in pulses per second or hertz (Hz) as shown in Figure 8. The PRF is related to the PRI by the equation:

$$PRF = \frac{1}{PRI} \quad (11)$$

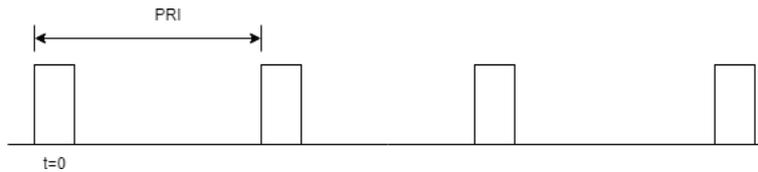


Figure 8. PRI in pulsed radar

The duty cycle (d_t) is the fraction of time the transmitter is on during one radar cycle, calculated as:

$$d_t = \frac{\tau}{PRI} = \tau \times PRF \quad (12)$$

where τ is the delay time. The average power (P_{avg}) of the transmitted EM wave is the product of the peak transmitted power (P_t) and the duty cycle (d_t) and PRF:

$$P_{avg} = P_t \times d_t \times PRF = \frac{P_t \times d_t}{PRI} \quad (13)$$

A comparative analysis of FMCW, Pulsed, and SFCW radars for target classification highlights the unique strengths and limitations of each system. FMCW radar provides superior range and velocity resolution, enabling continuous tracking of moving targets, making it ideal for real-time classification in dynamic environments, such as automotive and surveillance applications. Pulsed radar, while excelling in long-range detection and high-resolution profiling, requires more complex signal processing and higher power consumption, making it better suited for applications like military and aerospace. SFCW radar offers fine spatial resolution and performs well in noisy or cluttered environments but is less effective for tracking fast-moving targets. Given the need for both range and velocity resolution in most classification scenarios, FMCW radar is the most suitable choice for real-time, high-accuracy classification of moving targets.

3.4 Monostatic and bistatic radar

Radar systems typically use either a single antenna or closely placed antennas for transmission and reception, as shown in Figure 9. The transmitter emits an electromagnetic pulse that reflects off a target, and a duplexer switches the antenna to receiving mode to capture the echo signal. This co-located design simplifies radar systems, offering advantages such as reduced complexity, compactness, and cost-effectiveness due to fewer components. However, challenges include rapid switching between modes, which can cause interference, limit performance, and reduce range and sensitivity, limiting their ability to distinguish between closely spaced targets. In contrast, bi-static radar systems use separate antennas for transmission and reception, located at different sites, as illustrated in Figure 10. The spatial separation, ranging from a few meters to several kilometres, minimizes interference and enhances classification capabilities. This flexibility allows for optimized coverage and makes bi-static systems effective against stealth targets by complicating efforts to reduce radar cross-section as Magisetty et al. [31] demonstrates improved stealth detection using a nano-composite terpolymer in X-band radar.

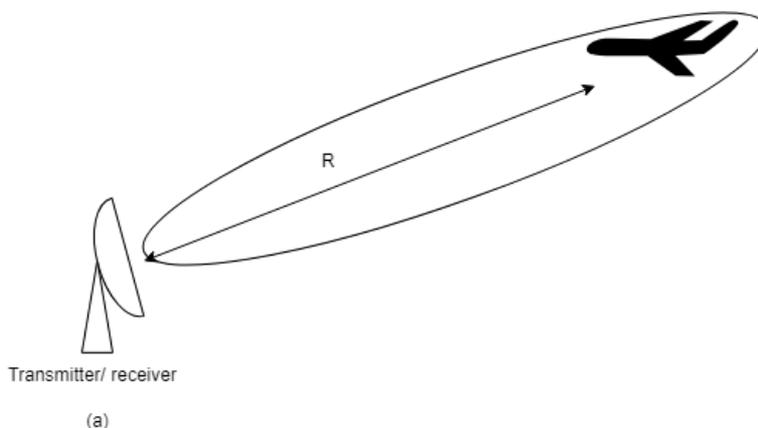


Figure 9. Monostatic radar

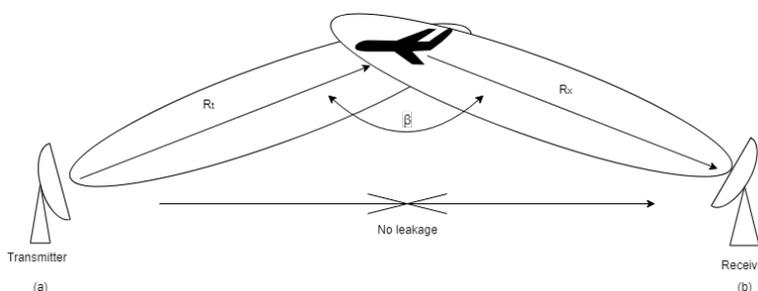


Figure 10. Bistatic radar

3.5 Non-coherent and coherent radars

Non-coherent radar systems detect only the amplitude of the received signal, disregarding phase information. These systems were foundational in early radar technology and are still used in applications where signal strength, rather than precision, is the primary concern, such as 2D ground mapping. In these cases, signal amplitude determines display brightness, helping visualize target environments. Non-coherent systems are advantageous due to their simplicity and cost-effectiveness, making them suitable for scenarios where targets can be easily distinguished from noise and high-resolution imaging is unnecessary. Historically, operators played a critical role in interpreting amplitude-based signals, making non-coherent radars pivotal in early radar operations. In contrast, coherent radar systems process both amplitude and phase information [32], treating the received signal as a vector in a complex plane. This approach enables precise analysis of target characteristics, capturing phase changes critical for determining velocity, direction, and motion. Coherent radars are integral in applications like Doppler radar for measuring target speed and synthetic aperture radar (SAR) for high-resolution imaging, where coherent processing [33, 34] produces detailed target images. These systems excel in target identification and clutter rejection by leveraging phase information to differentiate between targets and environmental noise. Consequently, coherent radars have become indispensable in modern radar technology, particularly in complex and dynamic environments.

With this foundation, the next section explains multiple classification methodologies. These techniques, ranging from traditional approaches to cutting edge machine/deep learning models, are designed to enhance the accuracy and efficiency of radar-based target classification.

4. Methodologies

In this section, various classification techniques employed for target classification are explored, starting with basic methods originally developed for classifying printed and handwritten text [35]. A range of approaches are examined, including traditional machine learning algorithms such as logistic regression and naïve bayes, as well as advanced deep learning techniques like convolutional and recurrent neural networks, self-attention mechanisms [36] and transformers. These methods are used to extract crucial features that are then utilized to train and optimize classification models, ultimately enhancing the accuracy and effectiveness of target classification in radar systems. All such techniques are elaborated below and summarized in the Table 9.

4.1 Using range FFT features

The methodology integrates a novel target classification technique utilizing mm-wave radars [37] and machine learning applied to a range of FFT features, alongside key procedural steps. It incorporates measurement setup and data acquisition [38], followed by the radar configuration and frame structure, and the application of FFT [39, 40] for signal analysis. While the IF signal provides a direct measure of object range, the range FFT plot undergoes statistical analysis—focusing on peak width, height, standard deviation and area, as shown in Figure 11—to extract distinctive features of various objects under test. Experiments classify common objects, including a pedestrian, car, and drones [41], using these features in conjunction with ML algorithms—Logistic Regression and Naive Bayes [42]. The subsequent steps in the methodology involve signal processing and feature extraction, leading to the deployment of ML models for object classification. This study uses machine learning on range FFT features for mm-wave radar target classification. Rigorous validation and performance evaluation of the proposed model is conducted.

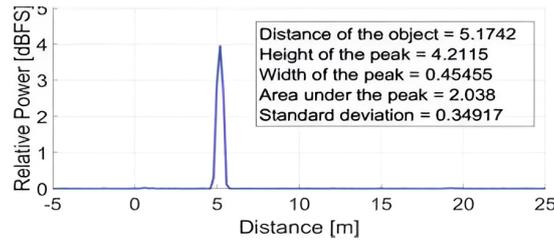


Figure 11. Features extracted from fast fourier transform

A. *Setup and Measurement*: A portable measurement setup was utilized, incorporating a complex baseband FMCW mm-Wave radar from Texas Instruments (TI), featuring 4 transmitters and 3 receivers, and operating within the 77–81 GHz frequency range. The objects tested included a compact drone, measuring $214 \times 91 \times 84$ mm when folded and $322 \times 242 \times 84$ mm when unfolded, a medium-sized car with dimensions of $4315 \times 1780 \times 1605$ mm, and a pedestrian represented by an adult with a height of 172 cm. A stationary radar unit was used to take measurements, and each object was moved up to a maximum distance of 25 m, which is the measuring setup’s limit. The FMCW signal pattern was used for data collection during these tests.

B. *Data Acquisition and Processing*: The data acquisition system employs a structure consisting of 4 distinct channels, with each channel encompassing a total of 200 frames. Within each frame, there are 128 chirps, and each chirp is sampled 256 times utilizing an Analog-to-Digital Converter (ADC). Consequently, the data is organized into a four-dimensional matrix with dimensions of $4 \times 200 \times 128 \times 256$. This arrangement facilitates the transmission of 128 chirps, ensuring that each chirp is accurately sampled 256 times, thereby enabling comprehensive data collection for all 128 chirps. Moreover, the echo signals, which are reflections of the transmitted chirps, are arranged in a similar fashion. This parallel organization results in the generation of two matrices of equivalent size, thereby allowing for efficient processing and analysis of both the transmitted signals and their corresponding echo responses. This structured data format is critical for subsequent signal processing and feature extraction, as it enables detailed examination of the radar’s performance in target detection and classification tasks. The Fast Fourier Transform (FFT) algorithm is employed to convert time-domain data into its frequency-domain representation by the following equation:

$$X_k = \sum_{m=0}^{n-1} x_m \cdot e^{\frac{j2\pi km}{n}} \quad (14)$$

Each chirp or frame is specifically subjected to FFT, which yields data that illustrates amplitude as a function of frequency. Using this frequency-domain data, a plot of amplitude (in decibel full scale, dBFS) vs range (in meters) is then produced, where the range is derived from frequency using a predefined equation. The dBFS value represents the full-scale amplitude of the signal. An algorithm for peak detection is subsequently applied to the range FFT plot. The algorithm looks for the maximum value stored in the matrix in every row; it corresponds to the beat frequency. The detected peaks within this plot correspond to the presence of targets within the azimuth field of view of the mm Wave radar. Upon identifying these peaks, key features are extracted for each detected peak [43, 44]. These features include the target’s radial range from the radar, standard deviation of the signal, as well as the height, width, and area under each peak. These extracted features are important inputs for the machine learning models used for target classification. Unlike conventional signal processing approaches that primarily rely on peak detection to ascertain the presence of a target within the radar’s field of view, this method leverages additional features, such as size and shape, to enhance classification accuracy.

By integrating these detailed features with lightweight machine learning models, the approach enables not only the detection but also the classification of various targets within the radar’s range. The dataset used for this study consists of three classes: human, car, and drone, with 95, 72, and 59 samples, respectively, resulting in a total of 226 data points. The dataset was divided into training and testing sets with a 9:1 split, allocating 23 samples for testing: 10 for humans, 7 for

cars, and 6 for drones. This dataset was used to train and evaluate two machine learning models: Logistic Regression and Naive Bayes.

Logistic Regression: A fundamental machine learning model that operates by applying two main steps. Firstly, it combines input characteristics and model parameters—such as weights and bias—in a linear fashion. Secondly, it applies a non-linear activation function, the sigmoid function, which generates a probability ‘p’ denoting the sample’s chance of falling into a specific class. Although Logistic Regression is typically used for binary classification [45], it is adapted it for multi-class classification to handle the three classes. This was achieved using the “one-vs-all” (or “one-vs-rest”) approach, where ‘n’ classifiers are constructed for ‘n’ classes. Until all classifiers have been generated, each classifier identifies one class as “0” and the other classes as “1” [46]. In this dataset, with $n = 3$, the Logistic Regression model was evaluated. The confusion matrix in Figure 12 indicates that the model predicted the ‘Car’ class with 0.86 accuracy, the ‘Drone’ class with 100 % accuracy, and the ‘Human’ class with 0.80 accuracy. The respective F1 scores, which represent the precision and recall weighted average, are 0.86, 0.92, and 0.84 for “Car”, “Drone”, and “Human”.

Naive Bayes: The Naive Bayes model, another machine learning algorithm, is also applied to the dataset. This model uses the Bayes theorem to calculate the likelihood that a given sample is a member of a specific class. Three variations of Naive Bayes exist, depending on the type of input features: Gaussian Naive Bayes for continuous features, Multinomial Naive Bayes for discrete features, and Bernoulli Naive Bayes for binary features. Given dataset consists of continuous features, the Gaussian Naive Bayes variant is used. The model first calculates the likelihood ratios for the dataset and then determines the posterior probabilities for each class. The sample is allocated to the class with the highest posterior probability. The evaluation of the Naive Bayes model showed that it predicted both the ‘Car’ and ‘Drone’ classes with 100% accuracy as shown in Figure 13, while the ‘Human’ class was predicted with 0.40 accuracy. The corresponding F1 scores are 0.88 for ‘Car’, 0.75 for ‘Drone’, and 0.57 for ‘Human’.

C. Performance Comparison: The performance of the Logistic Regression and Naive Bayes models is compared based on several metrics: accuracy, inference time, and model size. Logistic Regression achieved an accuracy of 0.869, with an inference time of 0.24 ms and a model size of 1 KB. The Naive Bayes model, on the other hand, attained an accuracy of 0.739, with an inference time of 0.1 ms and the same model size of 1 KB.

D. Limitations and Scope: This classification method avoids using complex machine learning models thus saving computational time and resources. When range is long and/or target cross section is small, then range FFT plot may not have distinguishable features and in such cases additional signal processing has to be explored before using machine learning. Incorporating micro-Doppler and range-Doppler features can improve classification for targets with vibrating parts or repeating patterns, making the model more robust.

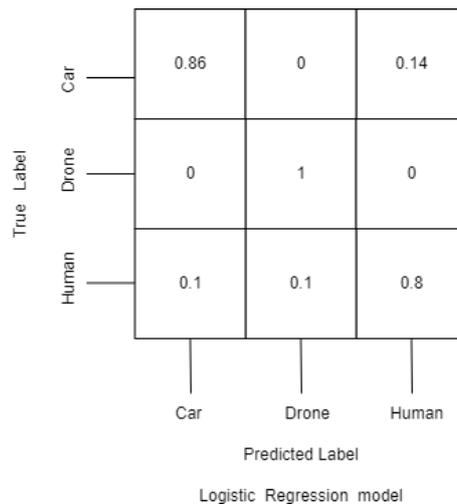


Figure 12. Confusion matrix for logistics regression model

| | | | | |
|------------|-------|-------------------|-------|-------|
| True Label | Car | 1 | 0 | 0 |
| | Drone | 0 | 1 | 0 |
| | Human | 0.2 | 0.4 | 0.4 |
| | | Car | Drone | Human |
| | | Predicted Label | | |
| | | Naive Bayes model | | |

Figure 13. Confusion matrix for naive bayes model

4.2 Using CNN on spectrogram

The FFT technique discussed by Muppala et al. [47] reduces the computational complexity to $\mathbf{O}((N \log \log N)^3)$. A huge amount of such high-resolution images, generated through this method, are fed to CNN architectures that learn the features via backpropagation and minimize the loss. Then trained models are deployed on the front for assistance [48]. Another method for classifying targets involves micro-Doppler signatures of the targets like humans [49]. This signature, received from FMCW radar [50] helps in creating the spectrogram of the target. The spectrogram provides a time-frequency representation of the radar signals reflected from the target. This information is crucial for analyzing the micro-Doppler signatures of the targets [51], which reveals information about their motion and characteristics [52]. This information is exploited to build real-time target detection and classification systems [53]. The classification techniques implemented classification algorithms like SVM [54, 55] or Bayesian Learning technique. The high complexity and large computation time of these algorithms make them unsuitable for implementation in real-time scenarios [56, 57]. To obtain the frequency-related information from the micro-Doppler signatures [58], STFT is carried out [59]. The micro-Doppler frequency is represented as:

$$f_g = \frac{2f}{c} [\omega \times r] \quad (15)$$

where f is the carrier frequency, ω is the angular velocity and r is the translational displacement. The STFT mathematical model is implemented to extract the micro-Doppler signal [60] component:

$$STFT\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t - \tau) e^{-i\omega t} dt \quad (16)$$

where τ is delay time. The spectrogram images are generated by obtaining the power spectral density function from this STFT which stores both the temporal and frequency information of the target. This is expressed as:

$$Spectrogram\{x(t)\}(\tau, \omega) \equiv |X(\tau, \omega)|^2 \quad (17)$$

The pre-existing CNN models like Google Net, Alex Net, and ResNet display higher accuracy on the data but take higher computation time. Therefore, Dr. Raj suggested a 15-layer CNN architecture, described in the Table 2, to train the 2400 bionic bird images and 2100 drone images over a 40-batch size with six iterations per epoch. This model classifies 200

images with an accuracy of 87.5% and takes much less time to produce results. Kim et al. [61] employed Google Net for drone classification, utilizing a pre-trained model and adapting it through transfer learning to meet their needs. Despite its high accuracy, Google Net is a deep network that requires significant time and computational resources for training on large datasets. The goal of this effort is to classify an ornithopter's behavior in safe locations, which requires a more accurate and computationally efficient network than Google Net. Zhang et al. [56] demonstrated that PCA, a resource-intensive machine learning technique, was used for feature extraction and classification. However, its high computational costs present a significant drawback compared to D-CNN approach. Another research conducted by Akella et al. [62] demonstrates superiority over these existing methods, successfully classifying different micro-Doppler signatures of various mini-UAV activities [63, 64, 65] such as static, flapping, and gliding ornithopters and quadcopters—with 97% accuracy utilizing radar STFT pictures. This network is suitable for deployment in secure areas to distinguish between ornithopters, birds, and quadcopters and requires fewer computational resources for training. In another research, Kumawat et al. [66] introduced DIAT- μ SAT dataset which consists of 6 classes, namely, Mini-helicopter, three short—and long-blade rotor, two-blade high-wind RC plane, quadcopter, and bionic bird, spread across 4849 micro-Doppler images. Here, STFT is performed using a hamming window of size of 256 samples, followed by fine-tuning the VGG-16 and VGG-19 architecture simultaneously for transfer learning [67]. This model achieves better scores for every class as shown in Table 3.

Table 2. 15-layer architecture of convolution neural network

| Layer | Name | Description |
|-------|---------------------------|--|
| 1 | Image input layer | $224 \times 224 \times 3$ image as input |
| 2 | Convolutional layer | $8 \times 3 \times 3$ Convolutions with 0 padding |
| 3 | Batch normalisation layer | Batch normalisation |
| 4 | ReLU layer | ReLU activation function layer |
| 5 | Max pooling layer | 2×2 max pooling with stride '2' |
| 6 | Convolutional layer | $16 \times 3 \times 3$ convolutions with 0 padding |
| 7 | Batch normalisation layer | Batch normalisation |
| 8 | ReLU layer | ReLU activation function layer |
| 9 | Max pooling layer | 2×2 max pooling with stride '2' |
| 10 | Convolutional layer | $64 \times 3 \times 3$ convolutions with 0 padding |
| 11 | Batch normalisation layer | Batch normalisation |
| 12 | ReLU layer | ReLU activation function layer |
| 13 | Fully connected layer | 4 fully connected layer |
| 14 | SoftMax layer | Softmax function layer |
| 15 | Classification layer | Final classification output |

Table 3. Performance of VGG16 and VGG19 on DIAT- μ SAT dataset

| Class | VGG 16 | | | VGG 19 | | |
|--------------------------------|-----------|--------|----------|-----------|--------|----------|
| | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| Two blades rotor | 0.93 | 0.96 | 0.94 | 0.97 | 0.95 | 0.96 |
| Three long blades rotor | 0.89 | 0.96 | 0.92 | 0.89 | 1.00 | 0.94 |
| Three short blades rotor | 0.96 | 0.89 | 0.92 | 0.97 | 0.91 | 0.94 |
| Bionic bird | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Two blades rotor & Bionic bird | 0.97 | 0.93 | 0.95 | 1.00 | 0.94 | 0.97 |
| Quadcopter | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 |
| Accuracy | | | 0.95 | | | 0.97 |
| Macro avg. | 0.96 | 0.95 | 0.95 | 0.97 | 0.97 | 0.97 |
| Weighted avg. | 0.96 | 0.95 | 0.95 | 0.97 | 0.97 | 0.97 |

Kumawat et al. [68] glorified the importance of small unmanned vehicle detection by processing the micro-Doppler signatures from the target. He proposed a RadSATNet architecture to accurately identify an RC plane, a three short-blade rotor, three long-blade rotor, quadcopter, bionic bird and mini-helicopter using X-band CW radar. Increasing the number of classes in a classification problem reduces the hallucination effect of the deep learning model. Having a greater number of classes than the previous models, this model shows better performance than other models discussed above. Moreover, restricting up to 256 samples saves spectral leakage, and improves frequency resolution and amplitude accuracy. Hence, it increases the computational efficiency as evident in the results.

4.3 Using residual learning in CNN

Bazil et al. [69] suggests a minimum identifying at least one component. The method includes selecting a minimal signature for one component that contains fewer components than a canonical signature. This technique does not involve regular CNN models trained over doppler signatures [70] which contain the characteristic features of each object. It uses FMCW radar to access frequency-intensity data within range-time domain to classify detected objects. Having mitigated the scarcity of the labeled data, this technique utilizes time stretching, frequency shifting and noise addition methods to further improve the model’s ability to generalize [71]. The CNN model, shown in Figure 14, is trained on raw radar data which helps in classification of targets: car, human and UAV. There’s been a great interest in the past few years for classifying objects by extracting features from the data. Nevertheless, various models such as Image Net [72, 73], Urban Sound 8k [74, 75] have made significant contributions in improving the accuracy of the overall performance in object classification, given that there exists a lot of data. The hidden relationship among features in range-time raw radar data is deeply understood by this model, which is sometimes not obvious to humans. Certain data augmentation techniques, such as time stretching, pitch shifting, and noise adding, compensate for the lack of data by somewhat improving performance.

A. CNN: This model slides over spectrogram representations to learn important features in each window. Such features are useful at other positions in the image [76]. The feature maps thus formed contain such features and other semantics. Max-pooling layer, which produces feature maps with maximum value within each kernel producing translational shifts within the given space tend to down-sample the features to avoid the curse of dimensionality. This reduces computational cost as well. For a given 2-D input X, kernel w, and cross correlation operation [77], is represented as:

$$y[i, j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x[m-i, n-j]w[m, n] \quad (18)$$

B. Residual Learning: Increase in depth of the layers gives rise to degradation problems, like gradient vanishing [78, 79], as it hampers the back propagation and introduces more parameters. It negatively affects the performance of the model. This leads to higher training and thus, test biases. To solve this problem, it is stated to use skip connection, as shown in Figure 15, which represents the underlying mapping that the network executes as $H(x)$ and the inputs to the first layer as x . In the event that, x is feed forwarded via a skip connection—which carries out identity mapping—the stacked layers will simply require learning residual function $F(x) = H(x) - x$. Because residual function optimizes more easily with this than with a conventional stacked network, performance loss is prevented. This enhances the model’s understanding of the underlying mappings. Also, it compensates the peripheral information lost due to convolution and pooling.

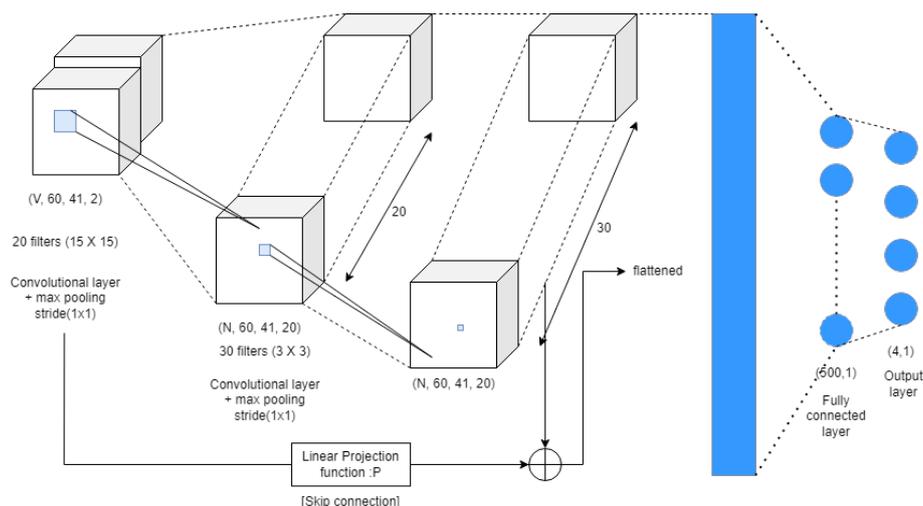


Figure 14. Double layer CNN architecture with skip connection

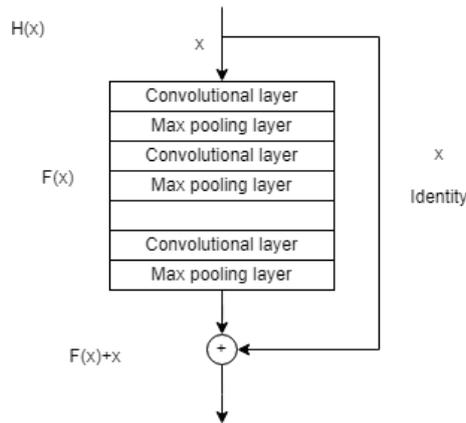


Figure 15. Block diagram of residual learning

C. *Data set*: The data set is collected through field experiments, of targets like, cars, humans and UAVs [80], moving within 30 m, the maximum range of the FMCW radar. The radar signal intensity data is collected in the range-time domain in the form of a wav file later to be pre-processed as a log-scaled Mel-spectrogram [81] using “librosa”. Instead of manually extracting features using MFCC (Mel-Frequency Cepstral Coefficients), CNN models are used to learn the underlying features and their relationships. Thereafter, data is augmented using:

Frequency Shifting: As shifting pitches in sounds don’t affect the harmonic relationships between different scales, in the same way, frequency shifting doesn’t affect the inter-frequency distances. It is applied to each sample’s data and sync channels. The time interval and amplitude of each data sample are maintained while shifting it by 4, 8, 16, and 32 half steps at a time.

Noise Addition: The noise w is created at random from the uniform distribution [2, 5], reshaped to fit the amplitude values of each sample’s length, multiplied by a scale c to regulate w ’s size, and added to the sample’s data channel while leaving the sync channel to retain its initial value. This translates the values along the amplitude axis.

Time Stretching: The raw data is converted into a frequency domain using a STFT and then a phase vocoder is used to time-stretch the signal by rate r . The data and sync channels are also changed by the augmentation as depicted in Figure 16.

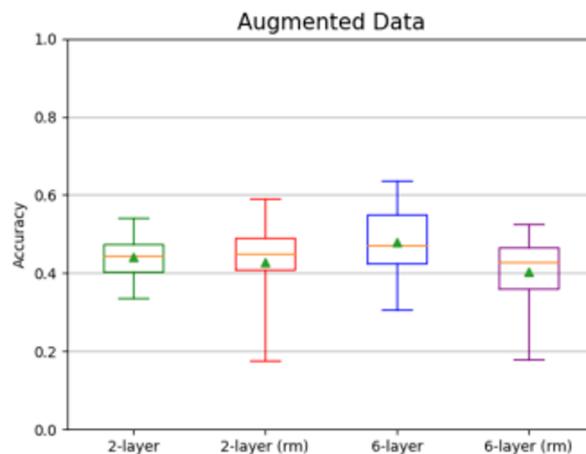


Figure 16. Augmented data distribution over layers

D. *Feature Extraction*: The raw range-time data shown in Figure 17 is sampled at 5682 Hz and then stored as a WAV file. After that, the Fast Fourier Transform (FFT) is used to transform each file into a frequency domain. This yields the

power spectrum, which is then multiplied by Mel-filters to create a Mel-scaled frequency. A log-scaled Mel-spectrogram is then obtained by applying these values with log. The window size that is being used is 512 (frames-1) and a hop length of 512. Mel bands are set to 60 with 41 frames. The original dataset consists of 1937 segments which is increased to 3944 by augmentation. Thereby generating segments of shape $M \times N$ where, M is the value of Mel-bands and N is the number of frames. The spectral features and their corresponding values are extracted using the “librosa” library. Each sample’s delta values are put together to create two input channels: delta and a log-scaled spectrogram.

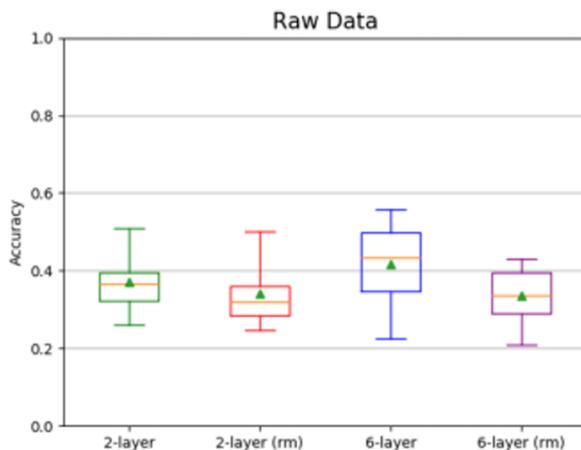


Figure 17. Simulated raw data distribution over layers

E. *Training Results*: The CNN model, from 2 to 8 layers, is trained using different sets of hyperparameters like depth of network, filter size, the number of filters, stride size, etc. The First Layer consists of 12 filters with a size (of 9×9) and 1×1 stride. The second Layer consists of 12 filters with a size of 5×5 and 1×1 stride. Both these layers are followed by ReLU and Max-Pooling with 9×9 pool size and 1×1 stride. 24 filters with a size of 5×5 and a stride of 1×1 make up the third layer. ReLU and Max-Pooling, which have a pool size of 5×5 and a stride of 1×1 , come next. The fourth Layer consists of 24 filters with size of 3×3 and stride 1×1 , followed by ReLU and Max-Pooling with pool size 5×5 and pool stride 1×1 .

Fifth and Sixth Layer consist of 48 filters each of size 3×3 and stride of 1×1 , followed by the ReLU and Max-Pooling of 3×3 pool size and 1×1 pool stride. 500 fully linked nodes make up the seventh layer, which is followed by a sigmoid activation function and Eight Layer consists of 4 output node which represents classes: human, car, UAV and others, followed by a *softmax* function. Each layer is first trained using the residual module and without it. The model is trained using mini-batch gradient descent of batch size 32. The training data is molded into 5 folds and only 4 folds are used for training and 1 for validation purposes. The best results are given by a 6-layer model without residual modules. The overall validation accuracy without residual is 47.96% and with residual is 40.44% and the per-class classification accuracy is 32.1% for UAV, 72.59% for humans and 43.58% for car and 36.37% for others is shown in Tables 4 and 5.

Table 4. Best layer accuracy evaluated on augmented data

| | Plain | Residual module |
|----------|--------|-----------------|
| 2-layers | 44.18% | 42.93% |
| 6-layers | 47.96% | 40.44% |

Table 5. Training results on targets

| | UAV | Human | Car | Others |
|--------------------|--------|--------|--------|--------|
| Per class accuracy | 32.10% | 72.59% | 43.58% | 36.37% |

F. Limitations and Scope: This method shows that that conventional data such as image and sound data are not the only type of data capable of holding unique, object-dependent features, but other seemingly featureless data such as range-time data can hold features that could be exploited by deep neural networks. Although classification accuracy of UAV seems relatively low, those of car and human show the possibility of using the raw radar data to train deep neural networks for classification. Since it is a DNN it uses significant number of computational resources. For higher accuracy it will require more computational resources. Still by using techniques like skip connection, residual learning performance is increased without increasing computational resources by much. since the dataset is in time-series, models that are effective for interpreting time-series data, such as RNN-based models, could unveil temporal dependencies among frequency values that could not be learned by CNNs.

4.4 Using CNN with GSAM

Shelly et al. [82] introduces an enhancement to the Alex Net model by incorporating an attention mechanism to improve the classification of human activity micro-Doppler signatures. The attention mechanism [83, 84] enables the model to focus on critical segments of the radar data, reducing the influence of background noise and enhancing prediction accuracy. The use of class activation mapping further elucidates the model's focus areas during predictions, thereby providing insights into its decision-making process. The attention mechanism is a strategic approach in deep learning that allows models to prioritize significant parts of the input data. It is akin to focusing on essential details in an image or text, enhancing the model's ability to discern crucial patterns and improve predictive accuracy.

A. Architecture: Alex Net is characterized by multiple layers, which autonomously detect patterns like edges, textures, or shapes in images. Micro-Doppler signatures of six participants of different heights, gender and ages groups performing eleven different activities of daily living is passed through eight layers: five convolutional layers for pattern detection and three fully connected layers for information integration, culminating in a classification layer that employs a softmax function to assign data to specific classes [85]. The initial two layers are succeeded by max-pooling layers, which streamlines size of the data and boosts model efficiency. Among Alex Net's innovations are the ReLU activation function, expediting learning, and dropout, mitigating overfitting. The Global Spatial Attention Module (GSAM) is employed to boost the model's sensitivity and predictive precision. GSAM generates a 2D spatial attention mask using feature maps from an intermediate layer. This mask accentuates vital regions by assigning higher values to pertinent areas while diminishing irrelevant details. GSAM integrates feature maps from the final convolutional layer with those from intermediate layers, allowing the model to concentrate on significant areas and disregard superfluous content. The global spatial attention map can be given by:

$$M_s(F^l) = \sigma_2 \left(f^{1 \times 1} \left(\sigma_1 \left(\widehat{F}_{Avg}^l + \widehat{G}_{Avg} \right) + \sigma_1 \left(\widehat{F}_{Max}^l + \widehat{G}_{Max} \right) \right) \right) \quad (19)$$

where, σ_1 and σ_2 are activation functions, $f^{1 \times 1}$ represents a convolution operation, \widehat{F}_{Avg}^l and \widehat{F}_{Max}^l are average-pooled and max-pooled features of intermediate feature maps F^l , and \widehat{G}_{Avg} and \widehat{G}_{Max} are pooled global features. In classic CNNs, feature maps are sequentially down-sampled to capture larger input areas, leading to varying spatial resolutions across layers. To formulate the attention mask, global feature maps are resized to correspond with the spatial dimensions of intermediate feature maps. The attention-refined features are computed through element-wise multiplication of the attention mask with down-sampled intermediate feature maps F^l , as shown below

$$\widetilde{F}^l = M_s \otimes \widehat{F}^l \quad (20)$$

where M_s is the attention map, and \hat{F}^l is the upsampled feature map to match the spatial resolution of the global feature map. The GSAM is integrated into the 2nd, 3rd, and 4th layers, enhancing local information capture. The 5th layer's activation maps are utilized for global feature extraction, maintaining spatial context crucial for radar micro-Doppler signatures.

B. Results: To compute attention-refined features, local feature maps from the 2nd, 3rd, and 4th layers, alongside global feature maps from the 5th layer shown in Figure 18, are processed through the GSAM, accentuating relevant regions in both local and global feature maps. Post-refinement, these maps are aggregated with global features to form the final prediction. Global average pooling is applied along spatial dimensions of refined and global feature maps, resulting in two vectors. These vectors are concatenated and passed through two fully connected layers to merge local and global data. The *softmax* function generates a probability distribution, with the class exhibiting the highest activation being selected as the final prediction. The proposed method effectively enhances Alex Net's performance in classifying radar micro-Doppler signatures, underscoring the advantage of integrating attention mechanisms in deep learning architectures.

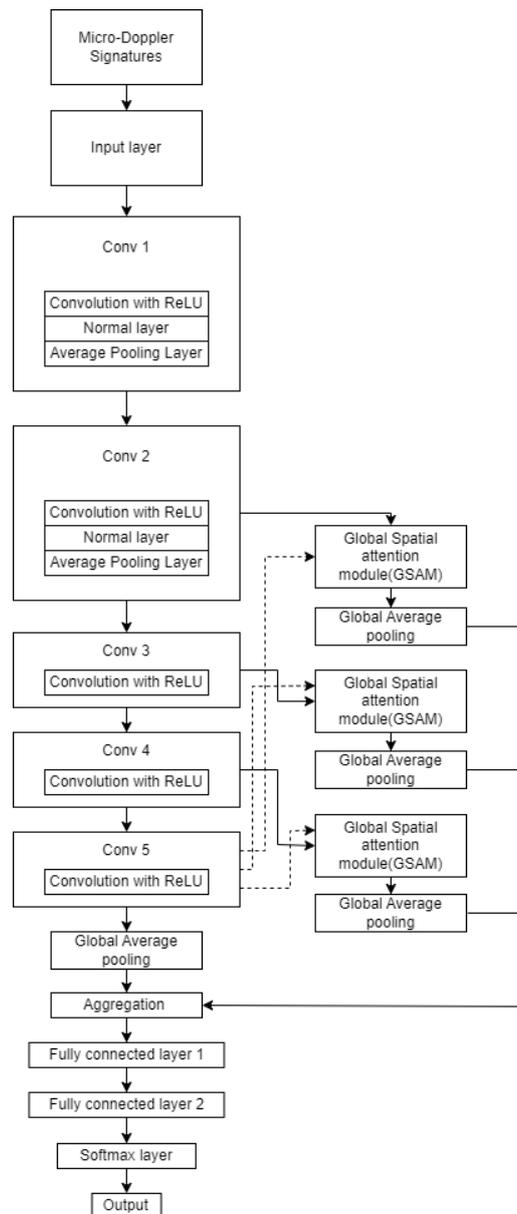


Figure 18. Attention enhanced Alex net architecture

4.5 Using RNN with self-attention

This technique [86] deals with the same problem and provides outstanding results. It uses self-attention mechanism shown in Figure 19 and gated recurrent unit, GRU [87] to process the sequence data. It is necessary to have enough knowledge about the surrounding signals in order to properly reconstruct the original signal. The test makes use of 3 different waveforms [88]: chirp sequence, triangular [89], and MFSK. These waveforms are targeted on the object, simultaneously, in k linear frequency chirps; The transmit signal's phase and frequency are as follows:

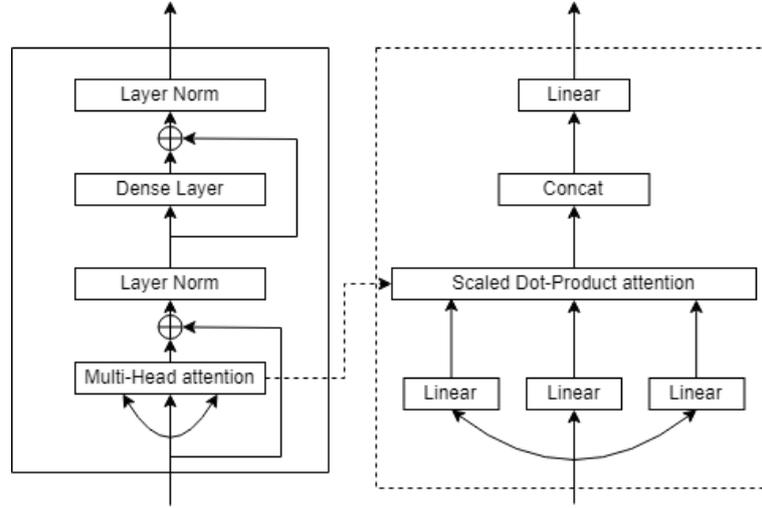


Figure 19. Architecture diagram of the self-attention model

$$f(t) = f_c + \alpha(t - kT_{chirp}) \quad (21)$$

$$\phi(t) = 2\pi \int_0^t f(t) dt = 2\pi(f_c t + \frac{1}{2} \alpha t^2 - \alpha k T_{chirp} t) \quad (22)$$

where T_{chirp} is the chirp time and B_{sw} is the sweep bandwidth. The FMCW waveform's slope is denoted by $\frac{B_{sw}}{T_{chirp}}$, while the transmitted signal's carrier frequency is represented by f_c . The difference between the transmitted and received frequencies is known as the beat frequency. Then it passes through an anti-aliasing filter which removes higher-frequency signals. Thus, the remaining beat phase is expressed as:

$$\phi_B(t) = \phi(t) - \phi(t - \tau) = 2\pi f_c \tau - \pi \alpha (\tau^2 - 2\tau t_k) \text{ if } \tau \leq t \leq T_{chirp} \quad (23)$$

The target's range R and velocity v is determined by applying 2D-FFT to the beat signal. The beat signal in the simulated environment seems like a smooth continuous wave but in real-time, a lot of atmospheric interference [90] and clutter creeps into it. Thus, rendering it of irregular shape.

A. *Preprocessing*: Firstly, the high-power interference is eliminated from the input by using a median filter. Depending on the transmitted signal's strength or signal attenuation from the distance between the transmitter and the target, the amplitude of the time samples of the beat signal changes significantly.

B. *Model*: The basic RNN cells suffer from a gradient vanishing problem when training on quite a deep network. In order to solve this problem, GRUs are being used in multiple layers as shown in Figure 20. The key to the attention block is Scaled-Dot-Product Attention. Its equation is expressed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (24)$$

where, Q , K , and V are the query, key, and value respectively and d_k is the dimension of the key vector. The loss function L used here is

$$L = \sum_{i=1}^T (y_{hati} - y_i)^2 \quad (25)$$

where y_{hati} is the label and y_i is the deep learning output. L is minimized using ADAM [91], a reliable and effective default optimizer for gradient descent algorithm. Assume that k out of n no. of channels are corrupted. The results for the model are shown in the Table 6.

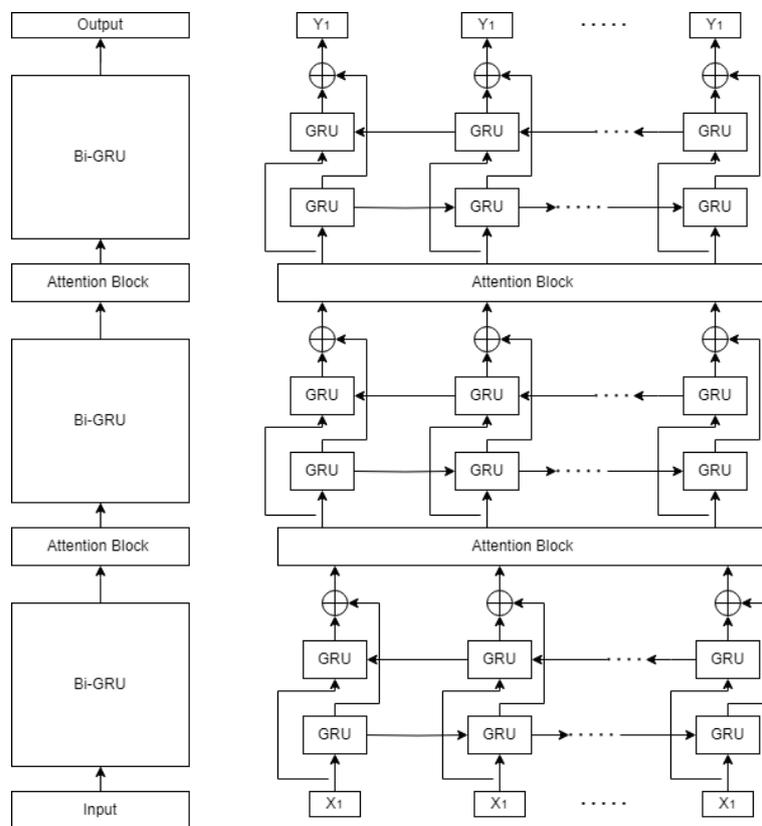


Figure 20. RNN architecture with attention block and gated recurrent unit

Table 6. Average SINR values for 50 trials

| | CS | Triangle FMCW | MFSK |
|------------------|--------|---------------|--------|
| No algorithm | 21.613 | 21.252 | 20.751 |
| Method I [8] | 25.415 | 25.038 | 25.076 |
| Method II [12] | 27.429 | 27.231 | 28.144 |
| Bi-RNN [20] | 33.187 | 32.254 | 33.920 |
| Attention Bi-RNN | 36.700 | 35.147 | 37.425 |

C. Limitations and Scope: The proposed method using RNN with self-attention effectively mitigates interference in FMCW and OFDM radar systems, outperforming traditional techniques by restoring original signals and improving SINR. Its advantages include enhanced accuracy, broad applicability, and better signal reconstruction, though it faces challenges like high computational demands, reliance on quality data, and limited real-world validation. This approach holds significant potential for autonomous driving, multi-radar environments, and real-time processing, offering a robust foundation for interference management in radar systems. The RNN with Self-Attention technique is highly effective for mitigating interference in FMCW and OFDM radars, outperforming traditional methods by reconstructing original signals and achieving superior SINR. Its ability to process complex temporal relationships makes it suitable for interference-prone environments. However, it is computationally intensive, relies on high-quality datasets, and requires real-world validation, limiting its immediate applicability in resource-constrained scenarios.

4.6 Using vision transformer

Micro-Doppler signatures, which reflect the micromotions of targets, provide crucial details about target characteristics. However, classifying low-resolution targets poses significant challenges due to limited spatial resolution and variations in frequency data. This study addresses these issues using the Vision Transformer (ViT), an advanced deep learning architecture known for effectively capturing global relationships across image patches. Unlike convolutional neural networks, which use convolutional filters to detect local spatial features, ViTs employ a self-attention mechanism [92] to model both local and global dependencies in the data as described by Liu et al. [93]. While CNNs excel at extracting localized features from high-resolution images, they are less effective at capturing long-range dependencies in low-resolution data like radar spectrograms. ViTs overcome this limitation by dividing the input spectrogram into patches and processing them as a sequence, similar to how words are analyzed in NLP tasks. This approach allows ViTs to comprehensively evaluate relationships between patches.

A. Signal Processing: Radar signals are processed within a Coherent Processing Interval (CPI) of 87 ms, during which multiple pulses are transmitted, and their echoes are accumulated. The relationship between CPI, the pulse repetition interval (PRI), and the number of pulses (N) is given by:

$$CPI\ Time = N \times cross\ PRI \quad (26)$$

Each radar pulse is represented in its in-phase (I) and quadrature (Q) components, expressed as:

$$x(n) = x_i(n) + jx_q(n) \quad (27)$$

The Doppler spectrum is then computed as:

$$X(f) = \sum_0^{N-1} x(n) \cdot \hat{e}(-\frac{j2\pi fn}{N}) \quad (28)$$

This spectrum provides an estimate of target velocity but lacks robustness against frequency fluctuations. To improve time-frequency distribution (TFD) analysis, the Short-Time Fourier Transform (STFT) is utilized. The STFT divides the radar signal into overlapping segments for localized frequency analysis. It is defined as:

$$STFT(f, t) = \sum_0^{L-1} w(l) \cdot x(l + t \cdot N_{step}) \cdot \hat{e}(-\frac{j2\pi fn}{N}) \quad (29)$$

where $w(l)$ is the window function of length L, here 128 and N_{step} is the step size. The STFT produces a spectrogram, which represents the energy distribution across time and frequency. The signal $x(n)$ of length N is divided into T, explain by the Figure 21 segments, given by:

$$T = \frac{N-L}{N_{step}} + 1 \quad (30)$$

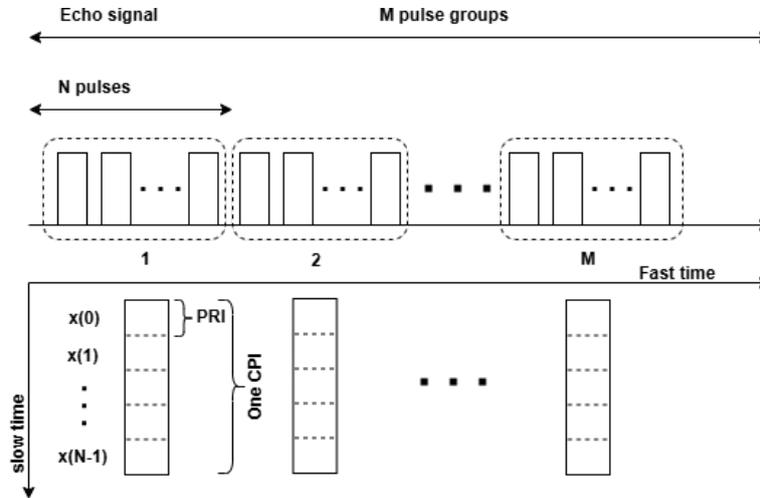


Figure 21. Coherent accumulation processing in fast and slow time domain

To address the loss of phase information in STFT, Bi-spectrum analysis is applied. It is defined as the Fourier Transform of the third-order cumulant sequence:

$$B_t(f_1, f_2) = STFT(f_1, t) \cdot STFT(f_2, t) \cdot STFT(f_1 + f_2, t) \quad (31)$$

The diagonal bi-spectrum, focusing on the main frequency components, is given by:

$$B_{diag}(f, t) = |B_t(f, f)| \quad (32)$$

By combining the STFT and Bi-spectrum outputs, a dual-channel spectrum is created, which incorporates both amplitude and phase information for enhanced analysis. Thus, forming 2 channel images.

B. *Architecture*: The Vision Transformer architecture processes input images through a series of well-defined steps, shown in the Figure 22, utilizing a self-attention mechanism for feature extraction and classification. The pre-processing is broken down into four main steps:

a. *Patch Embedding*: The first step involves partitioning the input image $I = H \times W \times C$, $H, W, C \in \mathbb{R}$ into non-overlapping patches of size $P \times P$. For an image with height H , width W , and channel depth C , this results in a total of $\left(\frac{H}{P}\right) \times \left(\frac{W}{P}\right)$ patches. Each patch is then flattened into a vector and passed through a linear projection layer, which maps it to a fixed-dimensional embedding space d . This is mathematically expressed as:

$$x_{patch}^i = W_p \cdot Flatten(I_{patch}^i) + b_p \quad (33)$$

where W_p and b_p represent the learnable weights and biases of the linear projection, and I_{patch}^i denotes the i -th image patch. To preserve the spatial relationships between patches, positional encodings are added to the embeddings, forming a sequence of inputs:

$$z_0 = [x_{patch}^1 + p^1; x_{patch}^2 + p^2; \dots; x_{patch}^N + p^N] \quad (34)$$

where p^i represents the positional encoding for the i -th patch, and $N = \left(\frac{H}{P}\right) \cdot \left(\frac{W}{P}\right)$ is the total number of patches.



Figure 22. Overall architecture of proposed approach

b. *Transformer Encoder*: The sequence of patch embeddings z_0 in the Vision Transformer is processed through a series of transformer encoder blocks. Each encoder block consists of two primary components: multi-head self-attention (MHSA) and a feed-forward neural network (FFN), both preceded by layer normalization. The MHSA mechanism computes relationships between patches by using the query Q , key K and value V matrices derived from the input embeddings. These matrices are calculated as:

$$Q = W_Q \cdot z, \quad K = W_K \cdot z \quad (35)$$

$$\text{and, } V = W_V \cdot z \quad (36)$$

where, W_Q , W_K and W_V are learnable weight matrices. z represents the input embeddings. The attention scores are computed using the following equation:

$$Attention(Q, K, V) = softmax\left(Q \frac{K^T}{\sqrt{d_k}}\right) V \quad (37)$$

where d_k is the dimensionality of the key vectors and the *softmax* function ensures the scores are normalized. This mechanism allows the model to capture dependencies between patches and focus on the most relevant regions of the input data. After the self-attention step, the output is passed through a position-wise FFN. The FFN consists of two linear layers, separated by a GELU (Gaussian Error Linear Unit) activation function:

$$FFN(x) = Linear_2(GELU(Linear_1(x))) \quad (38)$$

where the FFN introduces non-linearity and adds representational depth, enabling the model to capture complex patterns in the data. The output of each component is combined with the input using skip connections (residual connections), which help preserve gradient flow and stabilize training. The steps are:

$$\text{After MHSA: } z_{out} = LN(MHSA(z)) + z \quad (39)$$

$$\text{After FFN: } z_{out} = LN(FFN(z_{out})) + z_{out} \quad (40)$$

These skip connections ensure that the model effectively learns both shallow and deep features without vanishing gradients. The Transformer Encoder blocks repeat these operations in sequence, progressively refining the patch embeddings for feature extraction and eventual classification.

C. *Classification Token (CLS)*: A learnable classification token (CLS) is prepended to the sequence of patch embeddings at the beginning. This token serves as a global representative of the entire image. During self-attention operations, the CLS token interacts with all the patch embeddings, aggregating information from all patches. At the end of the transformer encoder stack, the CLS token absorbed the most relevant features from the input image, encapsulating its global representation. This token's role is crucial for downstream tasks, such as image classification, as it effectively summarizes the image in a single embedding.

D. *Classification Head*: The final step in the vision transformer architecture involves the classification head, which processes the output embedding of the CLS token. The embedding corresponding to the CLS token z_{CLS} is extracted after passing through the transformer encoder stack. The z_{CLS} embedding is passed through a fully connected layer, which computes a probability distribution over the target classes. The mathematical formulation is:

$$y = \text{softmax}(W_{CLS} \cdot z_{CLS} + b_{CLS}) \quad (41)$$

where, W_{CLS} are the learnable weights of the classification layer, b_{CLS} are the biases and z_{CLS} is the output embedding of the CLS token. The softmax function converts the scores into a normalized probability distribution, where each value represents the likelihood of the input image belonging to a specific class. By combining the classification token and the classification head, ViT effectively transforms input images into meaningful class predictions, leveraging the self-attention mechanism to process and summarize image information.

C. *Result*: The proposed method is evaluated using a dataset comprising 6400 radar frames, representing six distinct target categories: wheeled vehicles (1300 samples), tracked vehicles (1000 samples), persons walking (1100 samples), persons running (1000 samples), unmanned aerial vehicles (UAVs) (1400 samples), and ships (600 samples). The dataset was collected using an X-band surveillance radar system under diverse conditions, including road and hillside deployments, to ensure a variety of target and environmental scenarios. The radar system operated within a detection range of 100 m to 5 km, providing sufficient variability in target distance and echo characteristics. The performance of the model is represented in Tables 7 and 8.

Table 7. Accuracy and F1-Score of the method

| Class | Precision | Recall | F1-Score | Accuracy (%) |
|-----------------|-----------|--------|----------|--------------|
| Wheeled vehicle | 0.936 | 0.945 | 0.940 | 94.36 |
| Tracked vehicle | 0.928 | 0.943 | 0.936 | 94.33 |
| Person walking | 0.926 | 0.906 | 0.916 | 90.61 |
| Person running | 0.997 | 0.997 | 0.997 | 99.67 |
| UAV | 0.986 | 0.991 | 0.988 | 99.05 |
| Ship | 0.977 | 0.961 | 0.969 | 96.11 |
| Average | 0.958 | 0.957 | 0.958 | 95.69 |

Table 8. Comparison of classification accuracy with STFT- and Bi-spectrum- based ViT

| Class | STFT + ViT | Bispectrum + ViT | Proposed method |
|-----------------|------------|------------------|-----------------|
| Wheeled vehicle | 92.07 | 89.18 | 94.36 |
| Tracked vehicle | 90.66 | 84.79 | 94.33 |
| Person walking | 88.97 | 81.57 | 90.61 |
| Person running | 96.54 | 95.27 | 99.67 |
| UAV | 94.59 | 92.97 | 99.05 |
| Ship | 91.81 | 93.40 | 96.11 |
| Average | 92.44 | 89.53 | 95.69 |

4.7 Using 2D singularity power spectrum

It introduces an innovative methodology for SAR image analysis and target detection. This approach leverages the two-dimensional singularity power spectrum (2D-SPS) within the time-frequency domain to enhance the detection of weak targets, particularly in low SNR scenarios. By integrating the two-dimensional pseudo-Wigner-Ville distribution (2D-PWVD) [94] for time-frequency analysis and singularity power spectrum estimation [95], the proposed method effectively captures coupling features from both spatial and frequency domains. Experimental results indicate that this approach outperforms traditional fractal-based methods in target detection accuracy and robustness. SAR technology is widely recognized for its ability to generate high-resolution imagery under adverse environmental and operational conditions. However, conventional techniques often exhibit limitations in detecting weak targets, especially in complex scenarios with extremely low SNRs. While fractal-based methods have been employed to analyze SAR images, they frequently fall short in precisely characterizing power distribution within these datasets. To address these challenges, this study introduces the singularity power spectrum (SPS) as a refined metric capable of accurately capturing the power spectrum along the singularity exponent domain, thereby providing a robust characterization of SAR image features. The 2D-PWVD algorithm is used for time frequency analysis of the SAR image $I(x, r)$ is written as,

$$PW(x, r, f_x, f_r) = \sum_{k=1}^{N1} \sum_{l=1}^{N2} h(k, l) \cdot I(x+k, r+l) \cdot I^*(x-k, r-l) \cdot e^{-j2\pi(\frac{kf_x}{N1} + \frac{l f_r}{N2})} \quad (42)$$

where, (x, r) is the time component and (f_x, f_r) is the frequency component. The proposed methodology employs the 2D-PWVD to transform SAR images into four-dimensional arrays, uncovering intricate coupling information across spatial and frequency domains while effectively mitigating cross-term interference in multi-component signals. Subsequently, singularity power spectrum estimation is performed by calculating local singularity exponents for each pixel in the transformed image. These exponents are grouped into subsets, and the power distribution of these subsets is utilized to compute the 2D-SPS. This process yields a comprehensive and reliable representation of SAR image features. For target detection, the maximum SPS value at each pixel is extracted and utilized as a feature vector. This vector serves as the input to a detection framework, which employs threshold optimization through receiver operating characteristic (ROC) analysis. This approach ensures high sensitivity and specificity in discriminating targets from background noise. The proposed method, diagrammatically represented in Figure 23, is validated experimentally using SAR images obtained from the Terra SAR-X radar satellite, encompassing scenarios with and without weak ship targets. The results demonstrated a clear distinction between target regions and background areas, with singularity exponent values effectively characterizing the targets. The framework achieved near-perfect classification performance, with an area under the ROC curve (AUC) reaching 1.0, significantly surpassing the detection capabilities of conventional constant false alarm rate methods and other SPS-based approaches. Additionally, a detailed parameter sensitivity analysis highlighted the critical influence of sliding window size and subset count on detection efficacy.

In conclusion, the proposed SPS-SAR methodology represents a substantial advancement in SAR image processing, particularly in detecting weak targets within noisy and complex environments. This approach has demonstrated potential to improve the operational efficiency of radar-based surveillance systems. Future research will focus on optimizing the framework's parameters and integrating the method with deep learning architectures to enhance feature extraction and facilitate multi-target classification, further advancing the capabilities of SAR image analysis technologies.

4.8 Using RadHARNet architecture

Chakraborty et al. [96, 97] introduced an innovative approach to target classification utilizing mm-wave radar in conjunction with machine learning techniques presented in this study, with a focus on features derived from range-FFT. The preparation of a comprehensive dataset is initiated, where radar data from a diverse group of human subjects engaged in various activities across different distances and orientations is captured. Rigorous preprocessing is applied to this data, including filtering and transformation via STFT, resulting in detailed time-frequency maps. Significant features are extracted from these maps, which are subsequently used for the training and validation of machine learning models. The

DIAT-RadHARNet, a novel D-CNN architecture, is employed for the classification of these features. Advanced techniques such as depth-wise separable convolutions and multi-scale filtering are integrated into the model to optimize performance while minimizing computational load. Comprehensive performance evaluations of the proposed models are conducted, with dataset preparation, D-CNN architecture, and results discussed in the subsequent sections. In the preparation of the dataset, a diverse group of 30 human subjects, representing various genders, weights, and heights, are selected. These subjects are instructed to perform different suspicious activities at distances ranging from 10 m to 0.5 km from the radar. The activities are performed at seven distinct orientations (0° , $\pm 15^\circ$, $\pm 30^\circ$, and $\pm 45^\circ$) to simulate realistic conditions. Raw radar data is collected over 3 s under varying environmental conditions and processed using MATLAB. A 600-Hz low-pass filter is applied to remove high-frequency noise, while clutter from stationary and slow-moving objects is suppressed through digital filtering techniques. The filtered signals are then transformed using STFT, utilizing a 2048-length hamming window, 4096-point DFT, and 97.65% overlap. The resulting spectrograms, which reveal intraclass correlations in torso, hand, leg, and body movements, are organized by class, resulting in a total of 3780 samples—80% of which are allocated for training and 20% for validation.

The DIAT-RadHARNet architecture shown in Figure 24 is inspired by MobileNetV2, InceptionV3, Senet, and Blaze Face, and is designed around five core principles to enhance performance and efficiency. Feature map depth in convolution operations is optimized by weighting channels based on their significance. Depth-wise separable convolutions are employed to reduce the model's parameter count and computational load. Multiple filter sizes (1×1 , 3×3 , 5×5) are used in depth wise convolutions to balance global and local feature extraction. Different-sized kernels are applied to the same input tensor to handle variability in human activities, orientations, and distances. Global average pooling (GAP) layers replace fully connected layers to minimize trainable parameters and improve model efficiency. The DIAT-RadHARNet model is composed of 55 layers, including 13 separable convolutional layers, max-pooling, batch normalization, ReLU activation, dropout, dense layers, and softmax for classification. Spectrogram images, resized to $256 \times 256 \times 3$, are processed by the model, which also includes normalization and augmentation. The architecture is mathematically represented as:

$$DIAT - RadHARNet(x) = \beta(f^{FC}(f^D(f^G(f^{MSFCA_i}(f^M(f^B(\sigma \times (f^{C_i}(x * w_i + b_i)))))))))) \quad (43)$$

where f^{C_i} denotes the convolutional layers, f^B represents batch normalization, f^M is max-pooling, f^{MSFCA_i} indicates multi-scale filtering and channel attention, f^G refers to the GAP layer, f^D is dropout, and f^{FC} refers to fully-connected layers. The initial convolutional layer applies 96 filters of size 3×3 with a stride of 2, producing an output of $128 \times 128 \times 96$, which is processed through max-pooling and batch normalization.

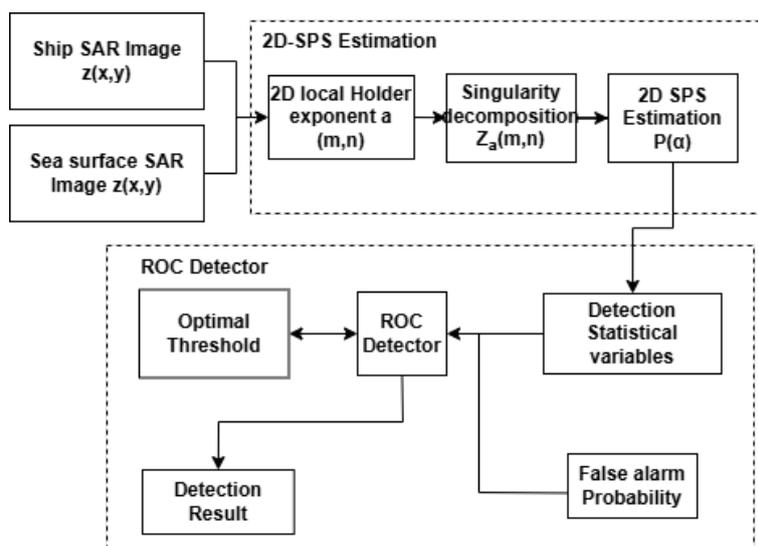


Figure 23. Flowchart of SAR ship detection based on 2D-SPS

The output is then fed into three Multi-Scale Filtering and Channel Attention (MSFCA) blocks, each containing convolutional layers and batch normalization. The output is fed into three Multi-Scale Filtering and Channel Attention (MSFCA) blocks. Each block includes:

$$f^{P1}(x) = f^B(\sigma(f^{C1}(f^x * w_i + b_i))) \quad (44)$$

$$f^{P2}(x) = f^B(\sigma(f^{C2}(f^x * w_i + b_i))) \quad (45)$$

$$f^{P3}(x) = f^B(\sigma(f^{C3}(f^x * w_i + b_i))) \quad (46)$$

where f^{C_i} is the convolutional layer, f^B is batch normalization, x is the input feature map, w_i and b_i represent weights and bias, respectively. The outputs from these paths are combined:

$$f^{A1}(x) = f^{P1}(x) \oplus f^{P2}(x) \oplus f^{P3}(x) \quad (47)$$

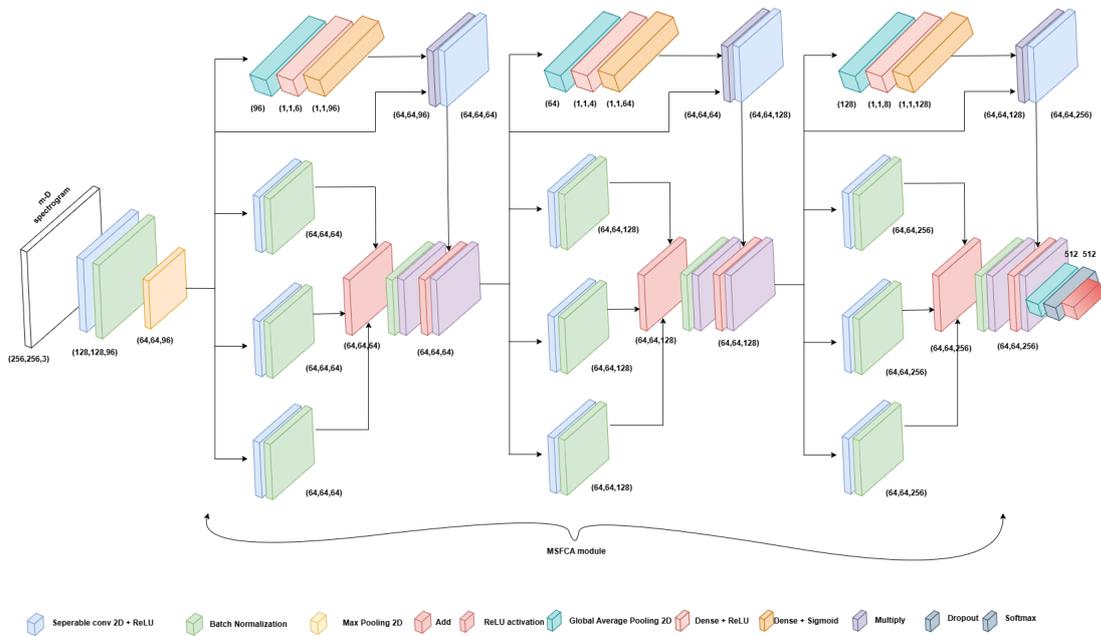


Figure 24. Layer description of the DIAT-RadHARNet architecture

The combined feature map $f^{A1}(x)$ undergoes further processing:

$$f^{A2}(x) = \sigma(f^B(f^{A1}(x))) \oplus f^{P4}(x) \quad (48)$$

$f^{P4}(x)$ captures channel interdependencies through GAP layer, followed by a multilayer perceptron that reduces the feature map to $1 \times 1 \times M$ and multiplies it with the original input:

$$f^{P_4}(x) = \sigma \left(f^{C_i} \left(f^{Mul} \left(\alpha \left(f^{FC} \left(\sigma \left(f^{FC} \left(f^G \left(f^x \right) \right) \right) \right) \right) \right) \right) \right) * f^x \quad (49)$$

where, $f^{P_4}(x)$ is the GAP layer, f^{C_i} represents the fully connected layers, f^{Mul} is the multiplication layer, and α is the sigmoid activation function. The outputs from these blocks are combined and further processed to capture channel interdependencies through GAP and a multilayer perceptron that reduces the feature map and multiplies it with the original input. Exceptional performance across varying environmental conditions was demonstrated by the DIAT-RadHARNet model. An accuracy of 99.22% was achieved under normal weather conditions, with only minimal misclassifications. High accuracy of 99.05% was maintained under adverse weather and low-light conditions, while the highest accuracy of 99.29% was observed during long-range target operations. Overall, the DIAT-RadHARNet model achieved a top accuracy of 99.22%, balancing high performance with computational efficiency. The model, comprising 213,793 parameters and 55 layers, is noted for its lightweight nature and superior performance compared to other models such as VGG-16, Mobile Net, and Blaze Face, providing an effective solution for radar-based target classification.

4.9 Using spatial frequency domain processing

Foliage Penetration (FOPEN) radar systems are critical for detecting and identifying objects disguised in dense foliage. To deal with foliage and environmental clutter, FOPEN radar systems use complex signal processing and imaging techniques. Polarimetric radar imaging [98] and synthetic aperture radar (SAR) are important tools for improving target recognition and imaging in such complicated situations. Range-Doppler methods are used to estimate target velocity, which helps identify moving objects hidden behind foliage. As Nakshatra et al. [99] illustrate, developing indigenous FOPEN radar technologies is critical for reducing reliance on imported systems and adapting solutions to unique operational settings. To maintain continuous performance, these systems must be flexible to changing weather conditions and foliage densities. Cognitive radar systems [100], which use machine learning and artificial intelligence, can augment FOPEN radar capabilities by allowing for autonomous decision-making and adaptive signal processing. FOPEN radar systems have applications other than ground-based target detection. They can also identify and track flying objects covered by tree canopies, enhancing their utility in security operations and environmental monitoring. The spatial frequency domain method represents a substantial advancement in the detection of moving targets using radar systems, addressing several limitations inherent in conventional Fourier transform techniques, such as difficulty in distinguishing target peaks and the requirement for prolonged time intervals to resolve all relevant frequencies accurately. These issues limit the ability for real-time processing and complicate radar data analysis. To overcome these limitations, the spatial frequency domain method offers a more refined and effective alternative. This approach involves transforming radar data into spatial frequencies [101], thereby enabling a more nuanced analysis of target detection. The transformation process entails converting each sample in the Pulse Repetition Period (PRP) into a spatial frequency, denoted as f_s . The relationship between spatial frequency and target distant R from radar is given by the formula:

$$f_s = \frac{2R}{c} \quad (50)$$

where c is the speed of light. This conversion is pivotal, as it allows the radar system to differentiate targets based on their distance from the radar. Targets located at greater distances generate higher spatial frequencies due to the longer round-trip travel time of the radar signal. Subsequent to the spatial frequency transformation, the IDFT is applied to the spatial frequency data. It is mathematically expressed as

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(f_k) e^{j2\pi \frac{kt}{N}} \quad (51)$$

where, $x(t)$ represents the time-domain signal, $X(f_k)$ denotes the spatial frequency components, N is the total number of samples, and t signifies time. Applying the IDFT [102], translates the spatial frequency data back into the time domain, which effectively reduces noise bandwidth and enhances the signal clarity. This improvement in clarity facilitates more

accurate identification of moving targets. A critical element of the spatial frequency domain method is clutter rejection. This process isolates moving targets from stationary clutter by comparing radar returns from successive pulses. Specifically, the difference between the current pulse return $R_{curr}(t)$ and the previous pulse return $R_{prev}(t)$ is computed using the formula

$$R_{diff}(t) = R_{curr}(t) - R_{prev}(t) \quad (52)$$

This subtraction effectively eliminates stationary targets, or clutter, leaving only the signals from moving objects. Consequently, the radar system can concentrate on analyzing the characteristics of these moving targets. The final step in the spatial frequency domain method involves analyzing target range and movement through RTI plotting. By generating an RTI plot from the IDFT results, the distance of targets over time is visualized. This plot provides two crucial pieces of information: the range, which indicates the distance of targets from the radar, and the intensity, which reflects the strength of the radar return signal and signifies the presence of targets.

5. Merits and demerits

Merits: FMCW radar systems are highly regarded for their exceptional resolution range and velocity measurement accuracy, making them ideal for precise target classification in applications such as autonomous driving and defence. This capability allows for effectively differentiating closely spaced targets in both distance and speed. Additionally, the continuous wave nature enables real-time data processing, which is crucial in dynamic environments. The system's robustness against interference and clutter further enhances its performance in noisy or complex environments, and its low cost and compact design make it versatile for various applications, from consumer electronics to industrial uses.

Table 9. Summary of the discussed techniques

| Reference | Radar signal representation | Network model | Task | Object type | Dataset | Remarks/Limitation |
|-------------------------|---|--|--|----------------------------|----------------|---|
| Bhatia et al. | Range FFT features | Logistic Regression, Naive Bayes | Target Classification | Cars, drones, humans | Simulated | Limited to basic feature extraction; struggles in cluttered or dynamic scenarios. |
| Upadhyay et al. | Spectrogram of micro-Doppler signatures | CNN | Motion Analysis and Classification | Moving objects | Self-developed | Requires large datasets; computationally expensive for real-time applications. |
| Vishwakarma et al. | Micro-Doppler signatures | CNN with Global Spatial Attention Module | Target Classification | Humans, vehicles | Custom Dataset | Improves noise resistance but increases computational complexity. |
| Han et al. | Raw radar data (range-time plots) | Residual CNN | Object Detection | UAVs, cars, humans | Simulated | Requires extensive preprocessing; reduced accuracy for UAV classification. |
| Mun et al. | Temporal radar signal data | RNN with Self-Attention | Interference Mitigation and Reconstruction | Moving and static targets | Simulated | High computational cost; limited real-world validation. |
| Ma et al. | Spectrogram patches | Vision Transformer | Classification | Low-resolution radar data | Public Dataset | Not suitable for real-time tasks without optimization. |
| Xiong et al. | SAR radar images | Custom Implementation | Weak Target Detection | Weak ships, static objects | Simulated | Computationally intensive; requires optimal parameter tuning. |
| Chakraborty et al. | Millimetre wave radar images | Custom Lightweight CNN | Human Activity Recognition | Humans | Custom Dataset | Limited application scope outside human activity scenarios. |
| Lakshminarayanan et al. | Spatial frequency representation | Custom Signal Processing Pipeline | Clutter Rejection and Target Detection | Moving targets in clutter | Simulated | Complex signal processing pipeline; struggles with dynamic, non-linear targets. |

Demerits: Despite its advantages, FM-CW radar also faces challenges, such as the need for complex and computationally intensive signal processing, which can increase system costs and complexity, particularly in real-time applications. The system's susceptibility to multipath effect can lead to inaccuracies in target classification, requiring advanced but not always effective signal processing techniques. Additionally, FM-CW radar performance can be heavily influenced by environmental conditions, such as weather and terrain, potentially reducing detection accuracy. Lastly, while it excels in range and velocity resolution, FM-CW radar often has limited angular resolution, making it difficult to distinguish between closely spaced targets with similar characteristics, which may necessitate more complex and costly antenna designs or additional sensors.

6. Future directions

The integration of advanced ML algorithms with FM-CW radar systems is anticipated to greatly enhance target classification accuracy. Future research is expected to focus on developing sophisticated signal processing methods to address current limitations like multipath effects, environmental dependencies, and signal noise. Hybrid sensor fusion, combining the radar with other modalities like LiDAR, cameras and ultrasonic sensors, is a promising approach to improving classification by leveraging complementary data. Moreover, ongoing efforts in miniaturization and cost reduction will be crucial for expanding the adoption of radar. The integration of real time AI, particularly through edge computing will further enhance the radar's ability to make quick, accurate decisions.

7. Conclusions

Radar attracted interest from AI community. The use of CNN with GSAM, RNN and self-attention, vision transformer, 2D-SPS are significant milestones because these methods solve the problem of micromotion detection of targets. These approaches are useful for reducing background noise and enhances prediction accuracy. The study highlights a range of methodologies, from traditional approaches to advanced machine and deep learning techniques, and assess their respective strengths and challenges. Radar systems advancement reviewed in Table 1 shows their robustness and ability to operate under diverse conditions, remain vital across industries such as, automotive, healthcare and security. Our review paper structured according to advanced radar design and mathematical model, including different types of radar and AI models. While recent advancements, such as the use of transformers, have significantly improved classification accuracy, they also introduce challenges like high data requirements and computational complexity. All these methods have associated advantages and disadvantages, since they are employed as an input to deep learning algorithms for various purposes. Traditional methods, though less demanding in resources, continue to offer viable solutions in specific scenarios. As radar evolves, addressing key challenges such as mitigating false positives and false negatives, improving object localization, and managing environmental complexities is essential.

Future advancements in radar systems will depend on improved hardware robustness and computational capabilities, enabling more efficient processing of complex datasets and enhancing image classification accuracy. Robust hardware will address environmental challenges, while advanced algorithms will tackle issues like false negatives, object localization, and handling multiple targets. These developments will improve precision and adaptability, ensuring radar technology continues to advance in various fields [103].

Conflict of interest

There is no conflict of interest for this study.

References

- [1] G. Sharma, A. B. Raj, "Effects of antenna radiation pattern on air borne SAR imaging with sufficient cyclic prefix OFDM," in *Proc. IEEE Conf. Inf. Commun. Technol.*, Chennai, India, Dec. 3–5, 2020, pp. 1-6.
- [2] G. Sharma, A. A. B. Raj, "Low complexity interference mitigation technique in IRCI Free SAR imaging algorithms," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, p. 4511305, 2022.
- [3] H. C. Kumawat, A. A. B. Raj, "SP-WVD with adaptive filter bank supported RF sensor for low RCS targets nonlinear micro-Doppler pattern imaging system," *Sensors*, vol. 22, no. 3, p. 1186, 2022.
- [4] A. M. Bork, "Physics Just Before Einstein: The development of electromagnetic theory was a major activity in physics before relativity theory," *Science*, vol. 152, pp. 597–603, 1966.
- [5] T. K. Sarkar, M. Salazar-Palma, and D. L. Sengupta, "Who Was James Clerk Maxwell and What Was and Is His Electromagnetic Theory?," *IEEE Antennas Propag. Mag.*, vol. 51, pp. 97–116, 2009.
- [6] D. K. Barton, "A half century of radar," *IEEE Trans. Microw. Theory Techn.*, vol. 32, pp. 1161–1170, 1984.
- [7] A. W. Price, "The evolution of electronic warfare equipment and techniques in the USA, 1901 to 1945," Ph.D. thesis, Loughb. Univ., 1985.
- [8] G. Marconi, "Wireless telegraphic communication," *Nobel Lecture*, vol. 11, pp. 198–222, 1909.
- [9] M. A. Tuve, "Radio Ranging and Nuclear Physics at The Carnegie Institution," in *Facets of Physics*, New York, NY, USA: Academic Press, 1970, pp. 163–177.
- [10] J. I. Glaser, "Fifty years of bistatic and multistatic radar," *IEE Proc. F*, vol. 133, no. 7, pp. 596–603, 1986.
- [11] J. L. Simmons and M. M. Backus, "A matched-filter approach to impedance estimation," *Geophysics*, vol. 61, pp. 484–495, 1996.
- [12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
- [13] P. Swerling, "Probability of Detection for Fluctuating Targets," *IEEE*, vol. 6, pp. 299–300, 1960.
- [14] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, 1965.
- [15] K. Ruttenberg and L. Chanzit, "High range resolution by means of pulse to pulse frequency shifting," *IEEE Eascon Record*, pp. 47–51, 1968.
- [16] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, 1986.
- [17] R. H. Roy III and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *Opt. Eng.*, vol. 29, pp. 296–313, 1990.
- [18] D. W. Bliss, K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution," *IEEE*, vol. 1, pp. 54–59, 2003.
- [19] A. Hassanien, S. A. Vorobyov, "Phased-MIMO radar: A tradeoff between phased-array and MIMO radars," *IEEE Trans. Signal Process.*, vol. 58, pp. 3137–3151, 2010.
- [20] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco, B. Himed, "Massive MIMO radar for target detection," *IEEE Trans. Signal Process.*, vol. 58, pp. 859–871, 2020.
- [21] A. Latpate, T. Gaikwad, M. Meshram, S. K. Patil, N. Akhter, and A. A. B. Raj, "A Review on Image Based classification for LSS Targets using Radar," *Int. J. Eng. Res. Rev.*, vol. 10, no. 4, pp. 100–112, 2022.
- [22] J. Farlik, M. Kratky, J. Casar, V. Stary, "Radar cross section and detection of small unmanned aerial vehicles," in *Proc. 2016 17th Int. Conf. Mech.-Mech. (ME)*, Prague, Czech Republic, Dec. 7–9, 2016, pp. 1-7.
- [23] B. S. De, "A survey on photonics technologies for radar applications," *J. Opt.*, vol. 52, no. 1, pp. 90–119, 2023.
- [24] R. B. Lakshmi P, "Design of 2D-WH/TS OCDMA PON ONU receiver with FLC technique," in *Proc. 2019 4th Int. Conf. Rec. Trends Elect. Info. Comm. Tech. (RTEICT)*, Bangalore, India, May 17–18, 2023.
- [25] V. D. Kanzarkar, K. S. Kayat, A. B. Jadhav, P. Tingare, V. P. Patil, D. E. Sampurna, et al., "Design of Digital Architecture for Real—Time T-F Imaging with New Radar Waveform," *Int. Res. J. Modernization Eng. Tech. Sci.*, vol. 5, no. 11, pp. 2380–2389, 2023.
- [26] N. A. Ansari, R. Kumar, A. Raj, P. Salotra, "Elevated levels of Ig3 and Ig4 subclass in paediatric cases of kala azar," *Parasite Immunol.*, vol. 30, no. 8, pp. 403–409, 2008.
- [27] B. R. N. Singh, "Estimation of heart rate and respiratory rate using imaging photoplethysmography technique," in *Proc. 2020 Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN)*, Pondicherry, India, Jul. 3–4, 2020.

- [28] C. L. Temes, "Relativistic consideration of Doppler shift," *IRE Trans. Aeronaut. Navig. Electron.*, vol. ANE-6, no. 1, p. 37, 1959.
- [29] M. T. Taba, S. M. H. Naghavi, M. Fayazi, A. Sadeghi, A. Cathelin, E. Afshari, "A compact CMOS 363 GHz autodyne FMCW radar with 57 GHz bandwidth for dental imaging," in *Proc. 2023 IEEE Cust. Int. Circ. Conf. (CICC)*, San Antonio, TX, USA, Apr. 23–26, 2023.
- [30] M. A. D. Mehta, "Stepped LFM Processing Methods for High Range," *Int. J. Elect. Electron. Data Commun.*, vol. 6, 2018.
- [31] R. P. Magisetty, A. B. Raj, S. Datar, A. Shukla, B. Kandasubramanian, "Nanocomposite engineered carbon fabric-mat as a passive metamaterial for stealth application," *J. Alloys Compd.*, vol. 848, p. 155771, 2020.
- [32] Q. He, R. S. Blum, "Non-coherent versus coherent MIMO radar: Performance and simplicity analysis," *Signal Process.*, vol. 92, pp. 2454-2463, 2012.
- [33] F. Gini, M. V. Greco, A. Farina, "Clairvoyant and adaptive signal detection in non-Gaussian clutter: a data dependent threshold interpretation," *IEEE Trans. Signal Process.*, vol. 47, pp. 1522-1531, 1999.
- [34] K. J. Sangston, K. R. Gerlach, "Coherent detection of radar targets in a non-Gaussian background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 330-340, 1994.
- [35] R. Srivastva, A. Raj, T. Patnaik, B. Kumar, "A survey on techniques of separation of machine learning printed text and handwritten text," *Int. J. Eng. Adv. Technol.*, vol. 2, no. 3, pp. 552-555, 2013.
- [36] A. Waswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, 2017.
- [37] C. Iovescu, S. Rao, "The fundamentals of millimeter wave sensors," *Texas Inst.*, pp. 1–8, 2017. Accessed: Oct. 15, 2024. [Online]. Available: <http://www.ee.fju.edu.tw/images/file/data3169/9a41b2657fe717233110268d676d38ef.pdf>
- [38] H. C. Kumawat, A. B. Raj, "Data acquisition and signal processing system for CW Radar," in *Proc. 2019 5th Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Pune, India, Sept. 19–21, 2019.
- [39] S. Björklund, H. Petersson, A. Nezirovic, M. B. Guldogan, F. Gustafsson, "Millimeter-wave radar micro-Doppler signatures of human motion," in *Proc. 2011 12th Int. Radar Symp. (IRS)*, Leipzig, Germany, Sept. 7–9, 2011.
- [40] D. Hand, P. Christen, "A note on F-measure for evaluating record linkage algorithms," *Stat. Comput.*, vol. 28, pp. 539-547, 2018.
- [41] D. K. Behera, A. B. Raj, "Drone detection and classification using deep learning," in *Proc. Int. Conf. Intell. Comput. Control Syst.*, Madurai, India, May 13–15, 2020.
- [42] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," In: R. Shoba, C. Mario, A. Mohammed, eds. *Encyclopedia of Bioinformatics and Computational Biology*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2019.
- [43] Y. Gao, M. T. Al Qaseer, R. Zoughi, "Complex permittivity extraction from synthetic aperture radar images," *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 4919-4929, 2019.
- [44] Y. Ding, G. Huang, J. Hu, Z. Li, J. Zhang, X. Liu, "Indoor target tracking using dual-frequency continuous-wave radar based on the range-only measurements," *IEEE Trans. Instrum. Meas.*, vol. 69, pp. 5385-5394, 2019.
- [45] D. Berger, "Introduction to binary logistic regression and propensity score analysis," *Res. Gate*, pp. 1–30, 2017. Accessed: Oct. 15, 2024. [Online]. Available: https://www.researchgate.net/publication/320505159_Introduction_to_Binary_Logistic_Regression_and_Propensity_Score_Analysis
- [46] R. Rifkin, A. Klautau, "In defence of one-vs-all classification," *J. Mach. Learn. Res.*, vol. 5, pp. 101-141, 2004.
- [47] A. V. Muppala, A. Y. Nashashibi, E. Afshari, K. Sarabandi, "Fast Fourier Time-Domain SAR Reconstruction for Millimeter-Wave FMCW 3-D Imaging," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 12, pp. 7028-7038, 2024.
- [48] B. Taha, A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669-138682, 2019.
- [49] S. Björklund, N. Wadströmer, "Target detection and classification of small drones by deep learning on radar micro-Doppler," in *Proc. 2019 Int. Radar Conf. (RADAR)*, Toulon, France, Sept. 23-27, 2019.
- [50] T. Y. Gite, P. G. Pradeep, A. A. B. Raj, "Design and evaluation of c-band FMCW radar system," in *Proc. 2018 2nd Int. Conf. Trends Elect. Informat. (ICOEI)*, Tirunelveli, India, May 11-12, 2018.
- [51] S. Rahman, D. A. Robertson, "Classification of drone and birds using convolutional neural networks applied to radar micro-Doppler spectrogram images," *IET Radar, Sonar Navig.*, vol. 14, pp. 653-661, 2020.
- [52] H. C. Kumawat, A. B. Raj, "Moving target detection in foliage environment using FMCW radar," in *Proc. 2020 5th Int. Conf. Commun. Elect. Syst. (ICCES)*, Coimbatore, India, Jun. 10-12, 2020.

- [53] M. Upadhyay, S. K. Murthy, A. B. Raj, "Intelligent System for Real time detection and classification of Aerial Targets Using CNN," in *Proc. 2021 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, May 6–8, 2021.
- [54] M. M. Hasan, M. Chakraborty, A. A. B. Raj, "A hyper-parameters tuned R-PCA+SVM technique for sUAV targets classification using the range micro-Doppler signatures," *IEEE Trans. Radar Syst.*, vol. 1, pp. 623–631, 2023.
- [55] M. L. Ch, A. A. B. Raj, L. Abhikshit, "Design and Implementation of a Secure Physical Unclonable Function In FPGA," in *Proc. 2020 2nd Int. Conf. Invent. Res. Comput. Appl. (ICIRCA)*, Coimbatore, India, Jul. 15-17, 2020.
- [56] P. Zhang, L. Yang, G. Chen, G. Li, "Classification of drones based on micro-Doppler signatures with dual band-radars sensors," in *Proc. 2017 Prog. Electromagn. Res. Symp.—Fall (PIERS—FALL)*, Singapore, Nov. 19-22, 2017.
- [57] A. B. Raj Anthonisamy, P. Durairaj, L. James Paul, "Performance analysis of free space optical communication in open-atmospheric turbulence conditions with beam wandering compensation control," *IET Commun.*, vol. 10, pp. 1096-1103, 2016.
- [58] R. D'Andrea, "Guest Editorial Can Drones Deliver?," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, pp. 647-648, 2014.
- [59] G. J. Mendis, T. Randeny, J. Wei, A. Madanayake, "Deep learning based doppler radar for micro UAS detection and classification," in *Proc. MILCOM 2016—2016 IEEE Mil. Commun. Conf.*, Baltimore, MD, USA, Nov. 1–3, 2016.
- [60] S. Rahman, D. Robertson, "Time-frequency analysis of millimeter-wave radar micro-Doppler data for small UAVs," in *Proc. 2017 Sens. Signal Process. Defence Conf. (SSPD)*, London, UK, Dec. 6–7, 2017.
- [61] B. K. Kim, H. S. Kang, S. O. Park, "Drone classification using convolutional neural networks with merged Doppler images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, pp. 38–42, 2016, DOI: 10.1109/LGRS.2016.2624820.
- [62] A. S. Akella, A. A. B. Raj, "DCNN based activity classification of ornithopter using radar micro-Doppler images," in *Proc. 2021 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, May 6–8, 2021.
- [63] S. Rahman, D. A. Robertson, "Radar micro-Doppler signatures of drones and birds at K-band and W-band," *Sci. Rep.*, vol. 8, no. 1, p. 17396, 2018.
- [64] J. Mazumder, A. B. Raj, "Detection and classification of UAV using propeller Doppler profiles for counter UAV Systems," in *Proc. 2020 5th Int. Conf. Commun. Elect. Syst. (ICCES)*, Coimbatore, India, Jun. 10–12, 2020.
- [65] M. M. Hasan, A. Pandey, A. A. B. Raj, "UAV Classification from Micro-Doppler Signature-based Time Frequency Images using SVM," in *Proc. 2022 Int. Conf. Augmented Intell. Sustain. Syst. (ICAISS)*, Trichy, India, Nov. 24–26, 2022.
- [66] H. C. Kumawat, M. Chakraborty, A. A. B. Raj, S. V. Dhavale, "DIAT- μ SAT: Small aerial targets' micro-Doppler signatures and their classification using CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1-5, 2021.
- [67] Y. Bengio, I. Goodfellow, A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [68] H. C. Kumawat, M. Chakraborty, A. A. B. Raj, "DIAT-RadSATNet-A novel lightweight DCNN architecture for micro-Doppler based SUAV targets' detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 71, p. 8504011, 2022.
- [69] A. Raj, P. Krothapalli, R. Mohindra, A. Sinha, P. Iyer, U.S. Patent Application No. 09/998,010, 2002.
- [70] D. P. Fairchild, R. M. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-Doppler Signatures," *IET Radar Sonar Navig.*, vol. 8, pp. 425-434, 2014.
- [71] H. Han, J. Kim, J. Park, Y. J. Lee, H. Jo, Y. Park, et al., "Object classification on raw radar data using convolutional neural networks," in *Proc. 2019 IEEE Sens. Appl. Symp. (SAS)*, Sophia Antipolis, France, Mar. 11–13, 2019.
- [72] A. Krizhevsky, I. Sutskever, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, pp. 84-90, 2017.
- [73] K. He, X. Zhang, S. Ren, J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. 2015 IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 7–13, 2015.
- [74] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *Proc. 2015 IEEE 25th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Boston, MA, USA, Sept. 17–20, 2015.
- [75] J. Salamon, J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Process. Lett.*, vol. 24, pp. 279-283, 2017.
- [76] G. E. Hinton, S. Sabour, N. Frosst, "Matrix capsules with EM routing," in *Proc. Int. Conf. Learning Repr.*, Vancouver, BC, Canada, Apr. 30–May 3, 2018.
- [77] B. M. Keel, J. M. Baden, T. H. Heath, "A comprehensive review of quasi-orthogonal waveforms," in *Proc. 2007 IEEE Radar Conf.*, Waltham, MA, USA, Apr. 17–20, 2007.

- [78] C. V. Cotton, D. P. W. Ellis, "Spectral vs spectro-temporal features for acoustic event detection," in *Proc. 2011 IEEE Work. Appl. Sig. Process. Aud. Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 16–19, 2011.
- [79] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 27–30, 2016.
- [80] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, I. Guvenc, "Micro-UAV detection and classification from RF fingerprints using machine learning techniques," in *Proc. 2019 IEEE Aeros. Conf.*, Big Sky, MT, USA, Mar. 2–9, 2019.
- [81] W. Jiang, Y. Ren, Y. Liu, J. Leng, "A method of radar target detection based on convolutional neural network," *Neural. Comput. Appl.*, vol. 33, no. 16, pp. 9835-9847, 2021.
- [82] S. Vishwakarma, W. Li, R. Adve, K. Chetty, "Learning Salient Features in Radar micro-Doppler Signatures using Attention Enhanced Alexnet," in *Proc. Int. Conf. on Radar Syst. (RADAR 2022)*, Hybrid Conference, Edinburgh, UK, Oct. 24–27, 2022, pp. 190-195.
- [83] D. Bahdanau, "Neural machine translation by jointly learning to align and translate," *arXiv*, 2014, arXiv:1409.0473
- [84] M. -T. Luong, "Effective approaches to attention-based neural machine translation," *arXiv*, 2015, arXiv:1508.04025.
- [85] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, "The history began from alexnet: A comprehensive survey on deep learning approaches," *arXiv*, 2018, arXiv:1803.01164.
- [86] J. Mun, S. Ha, J. Lee, "Automotive radar signal interference mitigation using RNN with self attention," in *Proc. 2020 IEEE Int. Conf. Acoust. Speech Sig. Process. (ICASSP)*, Barcelona, Spain, May 4–8, 2020.
- [87] J. Chung, C. Gulcehre, K. H. Cho, Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modelling," *arXiv*, 2014, arXiv:1412.3555.
- [88] J. R. Klauder, A. C. Price, S. Darlington, "The theory and design of chirp radars," *Bell Syst. Tech. J.*, vol. 39, pp. 745-808, 1960.
- [89] C. Sturm, W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communication and radar sensing," *Proc. IEEE*, vol. 99, pp. 1236-1259, 2011.
- [90] A. Arockia Basil Raj, S. Padmavathi, "Statistical analysis of accurate prediction of local atmospheric optical attenuation with a new model according to weather together with beamwandering compensation system: a season-wise experimental investigation," *J. Mod. Opt.*, vol. 63, no. 13, pp. 1286-1296, 2016.
- [91] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv*, 2014, arXiv:1412.6980.
- [92] B. Ma, K. O. Egiazarian, B. Chen, "Low-Resolution Radar Target Classification Using Vision Transformer Based on Micro-Doppler Signatures," *IEEE Sens. J.*, vol. 23, no. 22, pp. 28474-28485, 2023.
- [93] X. Liu, Y. Wu, W. Liang, Y. Cao, M. Li, "High resolution of SAR image classification using global-local network structure based on Vision Transformer and CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1-5, 2022.
- [94] J. Li, G. Xiong, B. Huang, "SAR Image Processing and Target Detection Based on 2D Singularity Power Spectrum in Time Frequency Domain," in *Proc. 2020 IEEE 5th Int. Conf. Sig. Image Process. (ICSIP)*, Nanjing, China, Oct. 23-25, 2020.
- [95] G. Xiong, F. Wang, L. Zhu, J. Li, W. Yu, "SAR Target Detection in Complex Scene Based on 2-D Singularity Power Spectrum Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9993-10003, 2019.
- [96] M. Chakraborty, H. C. Kumawat, S. V. Dhavale, "DIAT-RadHARNet: A lightweight DCNN for radar based classification of human suspicious activities," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-10, 2022.
- [97] M. Chakraborty, H. C. Kumawat, S. V. Dhavale, "Application of DNN for radar micro-Doppler signature-based human suspicious activity recognition," *Pattern Recognit. Lett.*, vol. 162, pp. 1-6, 2022.
- [98] J. M. Merlo, J. A. Nanzer, "A C-band fully polarimetric automotive synthetic aperture radar," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 2587-2600, 2021.
- [99] N. Bhavsar, A. Tayade, H. Ghodmare, N. M. Wagdharikar, N. Akhter, A. A. B. Raj, "A comprehensive review on foliage penetration radar systems," *Int. J. Eng. Res. Rev.*, vol. 10, pp. 28-40, 2022.
- [100] U. Garg, A. A. B. Raj, K. P. Ray, "Cognitive Radar Assisted Target Tracking: A Study," in *Proc. 2018 3rd Int. Conf. Commun. Electron. Syst. (ICCES)*, Coimbatore, India, Oct. 15-16, 2018.
- [101] P. Shakya, A. A. B. Raj, "Inverse Synthetic Aperture Radar Imaging Using Fourier Transform Technique," in *Proc. 2019 1st Int. Conf. Innov. Inf. Commun. Technol. (ICIICT)*, Chennai, India, Apr. 25-26, 2019.
- [102] O. Ponomareva, A. Ponomarev, V. Ponomarev, "Evolution of forward and inverse discrete Fourier transform," in *Proc. 2018 IEEE East-West Des. Test Symp. (EWDTS)*, Kazan, Russia, Sept. 14-17, 2018.
- [103] D. Yu, L. Deng, "Deep learning and its applications to signal and information processing [exploratory dsp]," *IEEE Sig. Process. Mag.*, vol. 28, pp. 145-154, 2010.