1. Introduction

Human action recognition (HAR) is a rapidly advancing field that focuses on identifying various activities performed by individuals based on observations of their actions and surrounding environment, through various sensor signals, including vision-based systems, wearable devices, and other sensor technologies like millimetre wave radar (mmWave radar) and, light detection and ranging (LiDAR) (Karim *et al.*, 2024). Traditional HAR methods primarily rely on camera-based systems, which, while effective, face significant challenges such as privacy concerns and limitations in varying lighting conditions (Karim *et al.*, 2024). These challenges have driven the exploration of alternative sensing technologies like millimetre-wave (mmWave) radar, which can penetrate obstacles and perform consistently across varying conditions (Huang *et al.*, 2023). Its large bandwidth, up to 4 GHz, complements the integration of mmWave radar in HAR and enables detailed detection capabilities in situations impacted by fog, smoke, or darkness. This makes it especially appropriate for non-intrusive monitoring in sensitive settings (Wu, Cui and Dahnoun, 2021). Furthermore, the synergy between mmWave radar and advanced artificial algorithms is enhancing HAR systems, enabling the integration of multiple data types, and depth information, to provide a better understanding of detected activities (Morshed *et al.*, 2023). While this capability is crucial for developing sophisticated systems that can accurately predict and analyse complex human actions, mmWave radar HAR systems face limitations such as rapid signal attenuation in air and vulnerability to background noise, which affect the accuracy and reliability of the data obtained (Cui and Dahnoun, 2021). This research seeks to enhance HAR's accuracy and reduce background noise, by employing advanced data processing techniques which include Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Kalman filtering, and the Hungarian Algorithm focusing on quantitative research methodologies to test hypotheses and provide answers to research questions using an existing dataset. The research initiates from a critical analysis of prior works, evaluating their methodologies, results, and scope explored, which are reviewed to set the foundation for defining the research trajectory, emphasizing the gaps and opportunities identified. Subsequently, the research adopts a structured project management approach to ensure systematic progression through the various phases of the methodology implementation. The practical phase involves the application of these data processing techniques for robust clustering, accurate prediction and noise reduction, and optimal assignment and tracking to test the feasibility and effectiveness of integrating these algorithms into HAR systems using mmWave radar individually as well as compatibility and enhancement potential when combined. This is followed by analysis with the use of existing model architectures with minimal tuning to validate enhancements, aiming to achieve superior recognition accuracy, which is crucial for demonstrating the practical applicability of the proposed methods. Finally, the research extends beyond technical validation to include a comprehensive business analysis, evaluating the real-world applications of this technology. This analysis considers business models to assess the commercial viability and potential market impact of the enhanced HAR systems, providing insights into their practical benefits and the economic opportunities they may offer.

1.1 Research aim and objectives

This research aims to develop an optimised processing workflow by combining DBSCAN, Kalman filtering, and Hungarian algorithm to enhance mmWave radar data quality for improved human action recognition through effective noise reduction and feature extraction. This is achieved through the following objectives:

1. Implementation of DBSCAN clustering for effective point cloud segmentation and noise reduction in mmWave radar data.
2. Integration of Hungarian algorithm for robust point tracking and interframe data association.
3. Application of Kalman filtering for trajectory smoothing and state estimation to detect human motion patterns.
4. Quantitatively evaluate the improvement in recognition accuracy and noise reduction achieved through the proposed algorithmic workflow compared to baseline processing.

1.2 Research Problems

Building on the preliminary insights discussed earlier regarding the potential and challenges of mmWave. radar, this research identifies a crucial problem area that demands further investigation: the effective reduction of environmental noise and signal interference in mmWave radar data, particularly addressing challenges such as rapid signal attenuation in air and background clutter, which significantly impact the accuracy of HAR systems. Current solutions struggle to differentiate between relevant human motion patterns and ambient noise, especially in dynamic environments with multiple moving objects. This problem is critical as noise can significantly hinder the system's ability to identify and track only the most relevant human-generated data clusters, which is essential for accurate and reliable activity recognition in diverse and dynamically changing environments. To tackle this issue, the research will leverage on the earlier mentioned data processing techniques, which are designed to meticulously filter out irrelevant noise and extract meaningful patterns from complex data. By focusing on this, the research seeks to advance the capabilities of mmWave radar-based HAR systems, making them more adaptable and efficient in real-world applications where precision, privacy, and reliability are paramount highlighting the transformative potential of mmWave radar technology in various business applications.

1.3 Business strategy analysis

From a business perspective, the development of mmWave radar-based HAR systems with enhanced noise reduction and motion detection capabilities opens significant opportunities for innovation and competitive advantage, particularly in privacy-sensitive monitoring applications. The virtual sensor market, which reached USD 8.78 billion in 2022, is projected to grow at a robust CAGR of 33.3%, reaching USD 15.4 billion by 2032 (DataHorizzon Research, 2023). As the virtual sensor market expands, it is reasonable to assume that the growth in computer vision-based HAR and related analytics will follow a similar trajectory, given that they both fall within the broader field of Artificial Intelligence (AI) (DataHorizzon Research, 2023).

To analyse the strategic implications of this technology, a SWOT analysis will be employed to identify the strengths, weaknesses, opportunities, and threats associated with the adoption of mmWave radar in Human Action Recognition (HAR). This analysis will offer valuable insights into how mmWave radar-based HAR systems can be effectively leveraged for commercial success, highlighting their potential in emerging market sectors.

2. Literature Review

This chapter provides a critical look at the current state of HAR systems. Starting by exploring the evolution of HAR, the background of mmWave radar, its role in human action recognition, and the advancements that have been made in real-time action recognition, in line with the research objectives outlined in the introductory chapter.

2.1 Human Action Recognition Background

HAR has evolved significantly over the years, which is influenced heavily by the integration of AI, and sensor fusion technologies. The field's foundations were built on computer vision, leveraging video data for activity recognition, and have since expanded to include radar-based and multimodal systems (Zeng, Zhang and Wang, 2020). This development from basic motion detection to complex action classification systems has made it possible to accurately identify human behaviours in real time for a variety of applications. mmWave radar has become a potent substitute for camera-based systems among these developments (Zeng, Zhang and Wang, 2020).

2.1.1 Early 2000s: Vision-Based Systems

The evolution of HAR can be traced back to the early 2000s when systems primarily relied on camera-based technologies for motion detection and activity recognition. These systems, while foundational, faced challenges such as dependence on ambient lighting, occlusions, and privacy concerns. To address these issues, techniques such as optical flow, spatiotemporal interest points (STIP), and background subtraction were used, but they struggled in complex scenarios (Hejazi and Abhayaratne, 2022). The introduction of Histogram of Oriented Gradients (HOG) and Motion History Images (MHI) improved feature extraction and motion understanding, but vision-based systems still faced difficulties in dynamic environments, particularly in multi-person scenarios (J. K. Aggarwal, 2011). A quantitative study by Hyukmin *et al.*, (2015) showed that MHI and HOG improved action recognition accuracy but still struggled in cluttered environments and with subtle actions.

2.1.2 Mid-2000s: Vision-Based Systems Enhanced by Machine Learning

The mid-2000s marked a critical transition with the adoption of machine learning techniques, further advancing computer vision methods. Support Vector Machines (SVMs) became a leading approach during this period, paired with advanced features like HOG and MHI. SVMs stood out for their ability to handle high-dimensional data and create optimal hyperplanes, making them well-suited for distinguishing between complex human activities (Wahyu, Agus and Nia, 2022)

In controlled settings, SVMs achieved up to 96% accuracy, as reported by (Wahyu, Agus and Nia, 2022) representing a substantial improvement over earlier methods. However, their limitations were clear. SVMs struggled with multi-class classification, particularly in noisy or cluttered environments where overlapping activities could confuse the model (Wahyu, Agus and Nia, 2022). Additionally, these systems depended heavily on large, labelled datasets, which were resource-intensive to generate and constrained scalability when applied to more complex or diverse action sequences. This pushed for the transition into radar systems.

2.1.3 2010s: Radar-Based Systems Through Deep Learning and Advanced Model Architectures

The 2010s marked a significant shift in HAR with the rise of deep learning, transforming both vision-based and radar systems for more accurate action recognition. Convolutional Neural Networks (CNNs) became central to this revolution, enabling automatic feature extraction from raw data and improving robustness against challenges like lighting, noise, and occlusions (Karim *et al.*, 2024). 2D CNNs processed spatial features, while 3D CNNs captured temporal dynamics, greatly benefiting applications like sports analysis and real-time monitoring (Michalis Vrigkas, 2015)

However, CNN-based systems faced challenges, including the need for large, annotated datasets and high computational demands, making real-time processing difficult. The performance also dropped in multi-person scenarios, with a decrease in accuracy when actions overlapped(Zeng, Zhang and Wang, 2020).

Radar systems emerged as an alternative, providing better performance in low-light or occluded conditions (Karim *et al.*, 2024b). While radar systems leveraged deep learning for micro-Doppler analysis, they still faced limitations in resolution and distinguishing fine-grained actions, highlighting the need for further advances in both sensor technology and algorithms (Singh *et al.*, 2021)(Roshan Singh a, 2023).

2.1.4 Current Era: Emergence of Unimodal and Multimodal Systems

Building on previous advancements, HAR systems evolved from unimodal approaches (single-input systems) to more sophisticated multimodal implementations. While unimodal systems benefited from refined architectures and transfer learning techniques to reduce dependency on large datasets they remained limited by environmental factors such as occlusions and lighting variations (Michalis Vrigkas, 2015).

The transition to multimodal systems, integrating multiple sensors (cameras, radar, depth sensors, and audio), significantly improved accuracy and robustness in challenging environments (Michalis Vrigkas, 2015; Bello, 2024). However, these systems face their own challenges, including complex data stream synchronization and high computational demands that affect real-time processing (Singh *et al.*, 2021). Multi-person tracking and overlapping action recognition remain particularly challenging.

This historical evolution demonstrates how computer vision, while historically dominant in HAR through camera-based methods, enabled high-resolution recognition of human activities in domains like surveillance and healthcare. However, vision-based systems are still limited by reliance on ambient lighting, vulnerability to occlusions, and privacy concerns in sensitive environments (J. K. Aggarwal, 2011). In contrast, mmWave radar presents a privacy-preserving alternative by capturing motion data without identifiable imagery. Its ability to penetrate obstacles and function effectively in low-light or occluded conditions makes it a compelling solution in scenarios where traditional vision-based systems fall short (Zeng, Zhang and Wang, 2020).

2.2 Millimetre-Wave Radar Technology

Building upon the evolution of HAR systems discussed earlier, mmWave radar emerges as an innovative sensing technology that addresses many limitations of traditional vision-based systems. Operating at frequencies between 30 GHz and 300 GHz, mmWave radar technology has evolved from its military origins to become a versatile tool in civilian applications, particularly in scenarios requiring non-intrusive, privacy-preserving monitoring (Al-Hourani *et al.*, 2018). This section explores the fundamental principles of mmWave radar technology, examining how it represents a significant advancement in the field of human action recognition.

2.2.1 Foundation of mmWave Radar

Historically, radar systems were primarily deployed in military and industrial applications due to their size, complexity, and cost. However, advancements in semiconductor technology have revolutionized radar systems, leading to compact, high-resolution mmWave radars for consumer level applications (Al-Hourani *et al.*, 2018). These advancements have unlocked significant research opportunities in HAR, where traditional methods, face challenges with privacy, lighting, and environmental variability. The radar offers unique capabilities such as high spatial resolution, material penetration, and robust performance across various environmental conditions (Appleby and Anderton, 2007; Al-Hourani *et al.*, 2018).

The foundation of mmWave radar lies in Frequency Modulated Continuous Wave (FMCW) radar, which simultaneously measures the range, velocity, and angle of targets. FMCW radars achieve this by transmitting frequency-modulated signals, called chirps, and analysing the reflected signals to determine key parameters (Liu, 2020). These features make mmWave radar particularly suitable for HAR, as it captures subtle human movements while preserving privacy and performing effectively in low-light or occluded conditions (Gu *et al.*, 2023).

Despite these advantages, mmWave radar is not without limitations. The technology's reliance on high-frequency signals provides excellent resolution but increases susceptibility to environmental noise, particularly in multi-person scenarios or areas with significant clutter. Additionally, the computational demands for real-time processing, especially for tasks like Angle of Arrival (AoA) estimation, pose challenges for widespread deployment in low-power or resource-constrained environments. The accuracy of AoA calculations, for instance, diminishes at larger angles, highlighting the need for further advancements in both hardware and signal processing techniques (Al-Hourani *et al.*, 2018).

2.2.2 Signal Processing in mmWave Radar

mmWave radar in HAR depends on signal processing methods that extract meaningful data from reflected signals. Three primary techniques are used which include the following.

i. Range Fourier Transform (Range-FFT)

Range FFT transforms the time domain intermediate frequency (IF) signals into the frequency domain to estimate the distance of targets. By analysing reflections, Range-FFT identifies discrete peaks in the frequency spectrum corresponding to targets at specific ranges (Iovescu and Rao, 2020). The distance ***d*** to a target is calculated using the following formula:

$$d = \frac{fIF^{C}}{2S}$$

Equation 1 - Range FFT

where:

- ***d***: Distance to the target,

- fIF: Frequency of the IF signal (beat frequency),
- c: Speed of light ($3 * 10^8\ m/s$),
- S: Chirp slope, defined as $S = \frac{B}{T_c}$,

  where $\boldsymbol{B}$ is the bandwidth and T_C is the chirp duration.

This method provides precise range detection, which is critical for isolating human subjects in crowded or cluttered environments (Huang *et al.*, 2023).

ii. Doppler Fourier Transform (Doppler-FFT)

Doppler-FFT estimates the relative velocity of targets by analysing phase variations between successive chirps (Zhao *et al.*, 2019; Ige and Noor, 2022). This technique is critical in HAR, as it generates unique motion signatures for activities like walking, falling, or running (Feng *et al.*, 2023). The velocity *v* of a target is determined using:

$$v = \frac{\lambda\omega}{4\pi T_c}$$

Equation 2 - Doppler FFT

where:

- v: Relative velocity of the target,
- λ: Wavelength of the transmitted signal,
- ω: Phase difference between the chirps,
- T_c : Time interval between successive chirps.

Doppler-FFT achieves velocity resolution as precise as 0.25 m/s, enabling the detection of subtle movement variations while maintaining privacy in HAR applications.

iii. Angle of Arrival (AoA) Estimation

AoA estimation uses multiple receiver antennas to determine the angular position of targets, which is crucial for tracking human movements in 3D space (Iovescu and Rao, 2020). This method is particularly valuable in scenarios requiring spatial awareness, such as multi-person tracking in crowded areas. AoA is calculated using phase differences between signals received by multiple antennas. For a two-antenna setup, the AoA θ given by:

$$\theta = sin^{-1}\left(\frac{\lambda\omega}{2\pi d}\right)$$

Equation 3- Angle of Arrival

where:

- θ: Angle of Arrival,
- λ: Wavelength of the transmitted signal,
- ω: Phase difference between received signals,
- d: Distance between the antennas.

By averaging results from multiple receiver pairs, AoA estimation achieves higher accuracy. However, its precision decreases for larger angles, which can be a challenge in certain HAR applications (Zhao *et al.*, 2019).

2.3 Advanced Processing Techniques

This research uses noise reduction techniques to enhance accuracy and robustness in HAR for mmWave radar signals.

2.3.1 Density-Based Clustering (DBSCAN) for robust clustering

DBSCAN is a foundational algorithm in unsupervised learning, recognised for its ability to identify clusters of arbitrary shapes while effectively managing noise in spatial datasets, which makes it a good fit for things like mmWave radar data, where clusters can be all sorts of shapes and sizes. It works by looking for areas where points are packed together tightly with the use epsilon (ε) and minsamples. Epsilon sets the maximum distance for points to be considered close to each other, which affects how big the clusters can be. minsamples sets the minimum number of points needed to form a dense cluster (Ahmed, 2016).

While DBSCAN is great at handling complex data, getting the best results means carefully choosing the right epsilon and minsamples values. One thing to keep in mind is that standard DBSCAN usually measures distance using a method that assumes data is spread out evenly. But with mmWave radar, that's not always the case movement might be more varied in one direction than another (Zhao *et al.*, 2019). This can make it harder for DBSCAN to accurately find clusters. Also, because of how it works, DBSCAN can be slow with very large datasets or when dealing with data that's constantly coming in. Researchers have been working on ways to improve this though, including using different ways of measuring distance that are better suited to radar data nature (Singh, Girdhar and Dahiya, 2022).

(Han *et al.*, 2023) conducted a comprehensive analysis of 3D point cloud descriptors in computer vision applications, particularly emphasising DBSCAN's clustering capabilities. Their study demonstrated DBSCAN's effectiveness in segmenting point clouds through its density-based approach, which proved superior to fixed grid-based methods, particularly in scenarios with significant point density variations. Results from this paper established that DBSCAN-based descriptors are effective for segmenting large and noisy 3D point clouds, offering better performance in scenarios with significant point density compared to fixed grid-based methods (Han *et al.*, 2023).

The limitations of the study indicate that while DBSCAN provides a solid foundation for point cloud segmentation, future research should focus on hybrid approaches combining DBSCAN with deep learning techniques to better handle complex, real-world scenarios (Han *et al.*, 2023).

These challenges suggest that while DBSCAN provides a solid foundation for point cloud segmentation, further refinement is necessary for more robust performance in complex real-world scenarios. Taking these limitations into account, this research proceeds by incorporating additional processing techniques, such as the Hungarian Algorithm, to improve data association and enhance clustering accuracy in dynamic and multi-target environments.

2.3.2 Hungarian Algorithm for Data Association

The Hungarian Algorithm is a combinatorial optimisation technique that solves the assignment problem, which involves pairing entities (e.g., detections and tracked objects) in a cost-minimizing manner. Originally introduced by (Kuhn, 1955), the algorithm operates on a cost matrix, where each entry represents the cost of associating a detection with a tracked object. Its primary strength lies in ensuring a one-to-one assignment that minimises the total cost across all pairings, guaranteeing optimal solutions in polynomial time. This makes it computationally efficient for real-time applications, a key factor in fields like human action recognition (HAR), particularly for multi-target tracking (Kuhn, 1955).

In HAR, the Hungarian Algorithm is effective for data association in dynamic environments like mmWave radar point clouds. It uses a cost matrix based on spatial and temporal features, such as distances between detections and predicted object positions, to ensure accurate associations, even in cases of overlapping trajectories or occlusions. Its optimality and computational efficiency make it ideal for real-time tracking in challenging radar conditions (Zhao *et al.*, 2019).

(Dang *et al.*, 2024) study leveraged the Hungarian algorithm as a key component in multi-person action recognition using mmwave radar data. The research aimed to address the fundamental challenge of maintaining consistent target tracking across frames while preserving privacy and handling complex environmental conditions.

The Hungarian algorithm was crucial in solving the data association problem after DBSCAN segmentation, identifying individual targets in each frame and maintaining their unique identities. It optimally matched targets between consecutive frames based on spatial and temporal relationships. (Dang *et al.*, 2024). The Hungarian algorithm demonstrated 92.2% classification accuracy in multi-person scenarios, proving its effectiveness in tracking targets. This performance validates its ability to handle complex multi-target tracking in radar data, overcoming issues like identity switches and track fragmentation, and its robustness in sparse and noisy radar point cloud data (Dang *et al.*, 2024).

However, the system showed limitations in scenarios involving rapid movements, where the Hungarian algorithm's frame-to-frame association might struggle to maintain tracking consistency (Dang *et al.*, 2024). This was particularly evident in highly dynamic environments where targets could exhibit sudden velocity changes or intersecting paths, challenging the algorithm's underlying assumption of minimal movement between frames(Dang *et al.*, 2024).

The Hungarian Algorithm is effective for consistent target tracking in complex radar environments, particularly in multi-person action recognition scenarios. However, its performance is limited in dynamic situations with rapid movements or significant velocity changes. Integrating with a filtering method can improve tracking accuracy.

2.3.3 Kalman Filtering for Smoothing and Prediction

Kalman filtering is a recursive estimation technique widely used to predict the state of dynamic systems from noisy measurements. Initially developed for aerospace applications, it has since become essential in many fields, including HAR, where it is used to track the motion of objects amidst sensor noise (Welch and Bishop, 1995). The Kalman filter works in two steps: prediction, where the system's next state is estimated based on its motion model, and correction, where the estimate is refined using new measurements. This iterative process allows it to effectively smooth radar data and predict future positions of tracked objects, making it invaluable for mmWave radar-based HAR systems, where noise from environmental factors like signal reflections can disrupt accuracy (Welch and Bishop, 1995; Chen, 2012) . The Kalman filter’s ability to balance sensor noise with system dynamics makes it highly effective in maintaining accurate, continuous tracking in dynamic environments.

(Pegoraro and Rossi, 2021) study developed a low-complexity, real-time system for tracking and identifying multiple subjects using sparse point-cloud data obtained from low-cost mm-wave radar with a focus on enhancing the accuracy and efficiency of tracking methods, particularly through the use of the Kalman filter.

The proposed system integrates the Kalman filter to estimate the position and shape of subjects, treating them as extended objects rather than point-shaped reflectors. This approach is combined with a novel deep learning classifier designed for effective feature extraction from radar point clouds. The system is implemented on an edge-computing platform (NVIDIA Jetson series) to evaluate its performance in real-time scenarios (Pegoraro and Rossi, 2021).

It achieved an accuracy of up to 91.62% while operating at 15 frames per second in identifying three subjects moving concurrently in an indoor environment. The Kalman filter demonstrated robustness in tracking trajectories even under complex and non-linear movements, effectively handling data association through a neural network-based logic (Pegoraro and Rossi, 2021).

A significant limitation identified in the study is the system's struggle with blockage events that can last several seconds, particularly as the number of subjects increases. This can lead to challenges in maintaining accurate tracking and identification during prolonged occlusions (Pegoraro and Rossi, 2021).

2.3.4 Comparative Studies

DBSCAN and K-means are commonly compared due to their differing clustering approaches: K-means is centroid-based, while DBSCAN is density-based. In the study by Andriyani and Puspitarani (2022), DBSCAN outperformed K-means in text clustering of product reviews, achieving 99.8% accuracy compared to K-means' 99.5%. DBSCAN's strength lies in handling noise and irregularly shaped clusters, but its performance can degrade in datasets with varying densities. K-means, while faster and more suitable for well-defined, evenly distributed data, struggle with noise and irregular shapes.

Kaushik and Kunjan (2013) compared both algorithms on time, cost, and performance, finding that DBSCAN excelled in clustering noisy and irregular data, while K-means performed better with simpler, well-separated data. DBSCAN's versatility is evident in real-world applications with noise and irregular shapes, whereas K-means is faster but less effective with complex data.

The Hungarian Algorithm and Auction Algorithm are compared for solving assignment problems. Arunachalam et al. (2020) found that while both are optimal, the Hungarian Algorithm is more computationally efficient, especially for larger problems. The Auction Algorithm, though useful in decentralized settings, faces delays with higher-density matrices.

Ismail and Sun (2017) showed that the Hungarian Algorithm outperforms the Auction Algorithm in scalability and convergence speed, making it ideal for centralized systems where computational efficiency is crucial. In contrast, the Auction Algorithm is better suited for decentralized scenarios, despite its slower convergence with larger problem sizes.

For Kalman filter, it being compared with extended Kalman filter (EKF), and particle filter (PF).

2.3.5 Integration and Justification of Selected Techniques

(Zhao *et al.*, 2019) developed a human tracking and identification system, mID, using millimeter-wave radar for non-intrusive applications like smart homes. The system integrates DBSCAN for clustering, Hungarian Algorithm for data association to ensure consistent tracking across frames, and Kalman Filtering for trajectory prediction of future positions and smoothen noisy data on generated sparse 3D point cloud data.

The system achieved 0.16 meters median tracking error and 89% identification accuracy for 12 people, performing well in cluttered environments with occlusions. However, scalability becomes a challenge as the number of people increases, and the radar struggles with sparse data and environmental interference. Despite these limitations, the integration of DBSCAN, Hungarian Algorithm, and Kalman Filter effectively addresses the key challenges of tracking and identification in dynamic environments, though further improvements in scalability and robustness are needed.

Building on the approach outlined in the paper, which successfully integrates all three algorithms, this research aims to justify the viability of this methodology for real-time applications. By adapting their methodology and using it as a benchmark, this study will apply the same algorithmic integration to a different dataset, allowing for comparative analysis and further exploration of its scalability and robustness in new environments. This will help refine the approach and address any limitations that emerge when applied to different dataset.

2.4 Research Gaps in mmWave Radar for HAR

Traditional vision-based HAR systems face fundamental limitations with lighting conditions, occlusions, and privacy concerns in sensitive environments. While mmWave radar technology offers promising solutions to these challenges, several critical research gaps remain in processing radar data for effective HAR (J. K. Aggarwal, 2011).

Current research struggles with extracting reliable features from noisy, sparse mmWave radar point clouds. Despite advances in radar technology, methods for handling varying point cloud densities across different human activities remain underdeveloped. The dynamic nature of human movement creates additional complexity in maintaining consistent feature representation across radar frames (Singh *et al.*, 2021).

Furthermore, they struggle to maintain reliable feature tracking across frames, particularly during complex, non-linear human movements. While some studies have explored basic tracking methods, comprehensive solutions combining clustering and temporal association remain limited (Dang *et al.*, 2024).

Most current studies focus on individual processing techniques rather than integrated approaches. There is limited research exploring optimised parameter tuning for combined clustering and tracking methods, particularly for real-time applications (Zhao *et al.*, 2019). This gap is especially evident in the context of processing mmWave radar data for HAR, where computational efficiency is crucial.

These research gaps highlight the need for integrated processing approaches that address the inherent challenges of mmWave radar-based HAR. Our proposed methodology, combining clustering, data association, and filtering techniques, aims to address these limitations through robust parameter optimization and comprehensive validation.

2.5 Significance of Study

The significance of this study lies in its exploration of mmWave radar-based Human Action Recognition (HAR) systems, focusing on overcoming the limitations of existing sensor technologies. By integrating techniques, the research aims to address challenges related to noise reduction, feature extraction, and data association. The findings will contribute to advancing the reliability, accuracy, and real-time performance of HAR systems, particularly in privacy-sensitive applications. This study also lays the groundwork for future research in optimising mmWave radar technology with advanced data processing.

3. Business Strategy Analysis

The Virtual Sensors Market, which includes mmWave radar, is experiencing significant growth, with projections for substantial expansion in the coming years as mentioned earlier. North America leads the market, with strong growth anticipated across industries such as oil and gas, automotive, healthcare, and manufacturing. The market encompasses both solutions and services, with deployment options like cloud and on-premises. Major players in the market include Honeywell, Siemens, and Texas Instruments, among others (DataHorizzon Research, 2023). Industry applications and analysis through a SWOT analysis are explored in this section.

3.1 Industry Applications and Use Cases of mmWave Radar in HAR

mmWave radar's unique capabilities make it an invaluable technology for Human Action Recognition (HAR) across various industries. Unlike camera-based systems, which rely on ambient lighting and struggle in low-light or occluded environments, mmWave radar operates effectively in total darkness and can penetrate obstacles like smoke or fog. This makes it ideal for critical industries such as emergency response and industrial settings, where traditional sensors fail to provide reliable data (Iovescu and Rao, 2020).

In healthcare, mmWave radar supports non-intrusive, privacy-preserving monitoring, and essential in-patient safety applications like fall detection and rehabilitation tracking, enabling real-time responsiveness and proactive care (Appleby and Anderton, 2007; Wang, Cang and Yu, 2019).

Within consumer technology, mmWave radar enhances wearables for monitoring daily activities and even facilitates inclusivity through real-time sign language recognition (Niemann *et al.*, 2022).

In industrial applications, mmWave radar's ability to offer context-aware HAR improves operational accuracy, such as in material handling within warehouses and distribution centres, streamlining processes and boosting efficiency (Reining *et al.*, 2019).

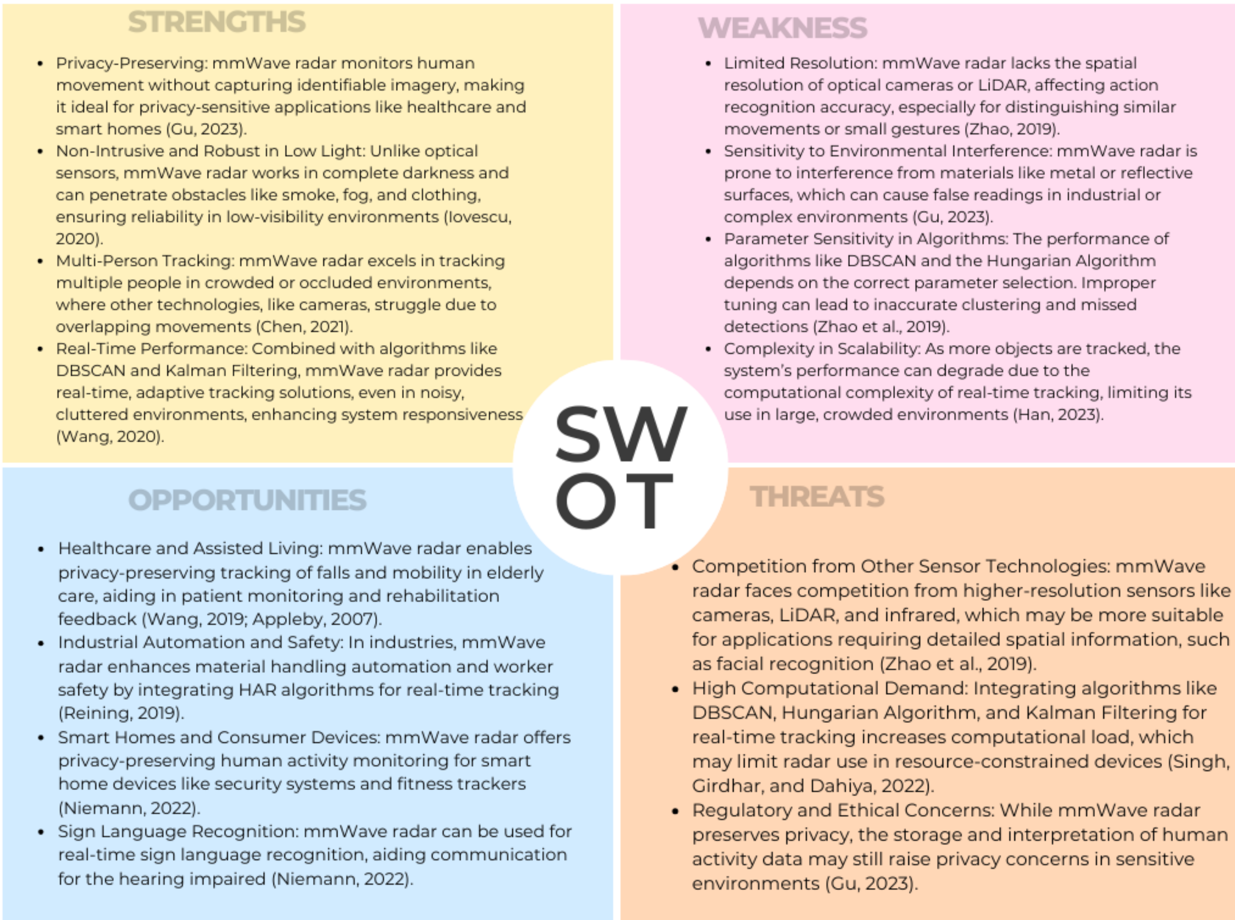
Overall, mmWave radar is transforming industry applications, offering robust, privacy-conscious, and real-time human activity recognition solutions that bridge technological innovation with practical industry needs.

3.2 SWOT Analysis

mmWave radar offers significant advantages for HAR, including privacy, strong performance in challenging conditions, and real-time application. However, it faces limitations like resolution constraints, sensitivity to interference, and scalability issues, necessitating further optimisation.

Continued research could make it a transformative technology in healthcare, industrial automation, and consumer devices.

Below is a detailed SWOT analysis in the picture.



4. Project Management

    4.1 Jira tracking

    4.2 Gantt chart

    4.3 Software Development/Techncal requiremnet

    4.4 Risk Assessment and Mitigation

    4.5 Workflow Overview (Flow chart)

5. Methodology and Implementation(All the components that I used)

This section begins by outlining the research questions and hypotheses that guide this study. These serve as the foundation for the implementation of the methodology and are systematically addressed through the subsequent evaluation presented in the following section.

5.1 Research questions

1. How effective is DBSCAN clustering in distinguishing between human subjects and similar obstacles in mmwave radar point cloud data?
2. How effective is the selection of the human cluster through the results of the kalman?
3. How do different DBSCAN parameters affect the clustering performance?
4. What are the optimal parameters for kalman filtering to maintain human?
5. What is the combined impact of all three techniques on the Milipoint dataset?

5.2 Research Hypotheses

1. The integration of DBSCAN clustering (independent variable) with mmWave radar data processing (independent variable) significantly improves the separation of noise clusters from human action patterns (dependent variable).
2. The application of the Kalman Filter combined with the Hungarian algorithm (independent variable) enhances the accuracy of action prediction and object identity tracking (dependent variables).
3. The noise reduction pipeline, incorporating DBSCAN clustering and Kalman-Hungarian tracking (independent variables), significantly improves the signal to noise ratio (dependent variable) in mmWave radar-based HAR systems.
4. The performance of mmWave radar-based HAR systems using the MiliPoint dataset (independent variable) is consistent in detecting and categorizing human actions (dependent variable).

5.3 Dataset Selection and Processing

5.2.1 Baseline Dataset Structure

The MiliPoint dataset (Cui and Yiren, 2023), was selected for this project due to its well-structured point cloud data, making it highly suitable for benchmarking clustering and tracking methods in HAR. It includes pre-segmented point cloud files, each representing specific human actions. These segments act as indirect activity labels, enabling a more streamlined analysis of clustering and tracking performance.

The dataset was designed to advance research in HAR, featuring data collected in controlled settings where participants performed a series of low-intensity, cardio-burning fitness movements. This approach resulted in a comprehensive compilation of 545,000 frames of point cloud data, contributed by 11 participants performing 49 distinct actions. Each frame is meticulously structured as a point cloud, with individual points represented in 3D space using X, Y, and Z coordinates (Cui *et al.*, 2023).

5.2.2 Data Preparation

This section describes the data preprocessing and transformation pipeline used to prepare raw data The pipeline was adapted from an existing repository to handle variability in point clouds and temporal sequences, ensuring consistency across frames.

<u>Overview of Classes</u>

- MMRKeypointData Class: This base class is designed to handle the loading and initial processing of 3D keypoints data from mmWave radar sensors. It provides essential functionalities like reading data from files, transforming it into a consistent format, and managing data files effectively.
- MMRActionData Class: Extending MMRKeypointData, this class adds functionalities specifically for action recognition tasks. It incorporates action labels into the dataset, enabling the training of supervised models. It also adjusts the data processing to cater specifically to the dynamics of action recognition by ensuring that temporal relationships within the data are maintained and appropriately labelled.

<u>Executed Data Preparation Pipeline</u>

- Data Loading and Configuration: Raw data was loaded from the specified directories as provided by the repository.
- Keypoint Transformation: Dimension Reduction: Data originally with 18 key points was often simplified to a 9-keypoint format by averaging certain key points, focusing on essential body parts relevant to the studied actions.
- Stacking and Padding:
    - Temporal Stacking: The stacks parameter determined how many consecutive frames were stacked together, capturing the necessary temporal sequence of movements for action recognition.

- Padding for Uniformity: Frames were padded with zeros to meet a predefined point limit, ensuring uniform input sizes across the dataset.
- Handling Large Frame Sizes: Stacked frames were configured to reach up to 1100 points, providing a dense and comprehensive representation of movement sequences.

- Data Partitioning and Management:
  - Dataset Division: The dataset was split into training, validation, and testing sets with proportions of 80%, 10%, and 10%, respectively.
  - Data Handling: The system efficiently managed data files, loading existing pre processed data when available to save processing time, or reprocessing data when necessary.

Appendix A shows the code for the data processing which resulted in the following characteristics;

- Total number of frames: 170336
- Structure of a frame: (1100, 3)

5.4 Overview of techniques used

5.2.1 DBSCAN clustering

DBSCAN uses two main parameters: epsilon (eps) and minsample. Epsilon defines the radius of a point's neighbourhood, where setting it too high or too low can improperly merge separate clusters or fragment them. Minsample sets the minimum number of points required in an eps radius to define a core point, influencing the cluster's density and the identification of noise. Adjusting these parameters carefully is essential for accurately capturing the underlying cluster structures, especially in complex or noisy data environments (Monko and Kimura, 2023).

i. Parameter selection

This section shows how all the DBSCAN parameters were selected

1. Min Sample

The minsample was set using the rule of thumb $\mathbf{minsapmle} \geq \boldsymbol{dimensionality} * \mathbf{2}$. With (x,y, z) spatial coordinates indicating 3 dimension, minsample was set to 6.This ensures a balance between detecting small clusters and filtering noise, as supported by (Monko and Kimura, 2023).

<u>Impact of minsample Choices</u>

- Low minsample (1): DBSCAN finds 9 clusters and no noise points. This suggests that the algorithm is very sensitive and considers even individual points as clusters. This might lead to overfitting, as very small groups or isolated points are treated as distinct clusters.

- Optimal minsample (6): DBSCAN finds 7 clusters and 2 noise points. This is a more balanced outcome, where the algorithm groups point into meaningful clusters and still identifies a few as noise. It shows the algorithm considers the minimum number of points for a group to be a cluster, leading to a reasonable balance between clustering and noise.
- High minsample (9): DBSCAN finds 7 clusters and 8 noise points. Increasing the minimum number of points required for a cluster result in more points being treated as noise, especially in sparser regions.

  Figure _ below shows the impact of the low, high, and optimal minsamples

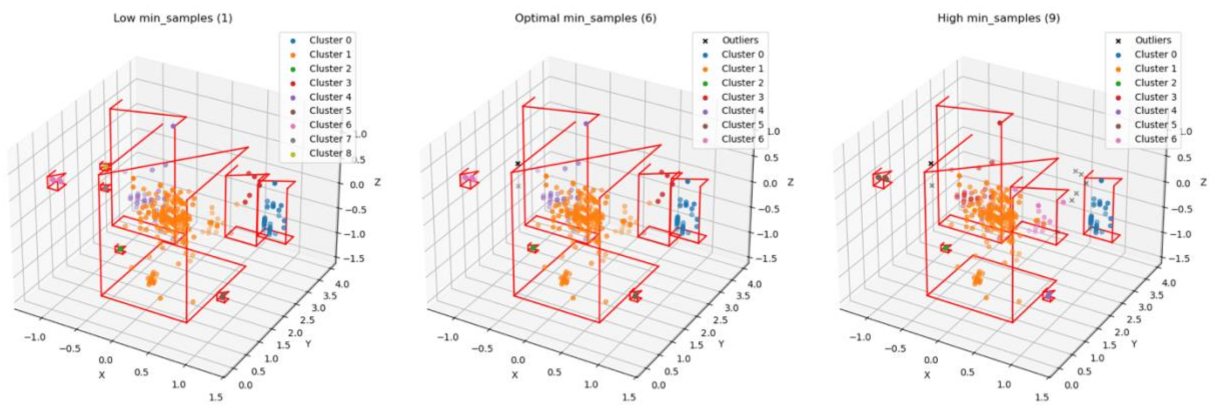


Figure 1 - Comparison of minsample values with low, optimal and high values

2. Silhouette Analysis

A grid search was conducted for eps values, evaluating clustering quality using the silhouette score. This metric measures how well points within a cluster are grouped and how distinct each cluster is.

Setup: eps in the range [0.1, 1.0] with steps of 0.1.

The result was plotted to visualise the silhouette score for different combinations of these parameters and is shown in fig. Which shows the optimal value to be **0.4** with a silhouette score of **0.8676**.

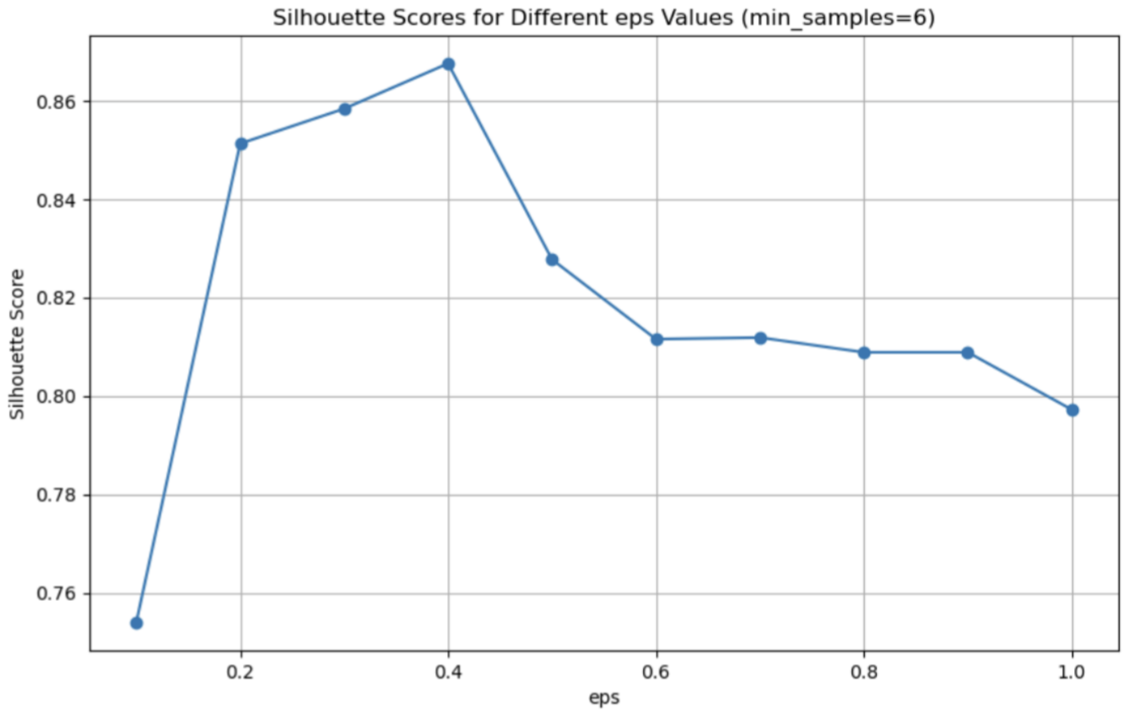


Figure 2 - Silhouette Scores for different eps values

<u>Impact of eps Choices</u>

- Low eps (0.2): With a low epsilon value, DBSCAN tends to generate more clusters, as seen with 9 clusters. However, this also results in a high number of noise points (23), as the algorithm is more sensitive to small changes and treats many points as outliers.
- Optimal eps (0.4): This epsilon value balances between creating a reasonable number of clusters (7 clusters) and minimizing noise points (only 2). It appears to be the optimal choice, as it produces sufficient noise points and still captures the overall structure of the data effectively.
- High eps (0.8): With a high epsilon value, DBSCAN merges more points into fewer clusters. This lead to only 2 clusters and no noise points, suggesting that the algorithm is less sensitive to finer structures in the data and treats more points as part of the same group. However, this can also result in losing finer details and distinct groups within the data.
  Figure _ below shows the impact of the low, high and the optimal eps

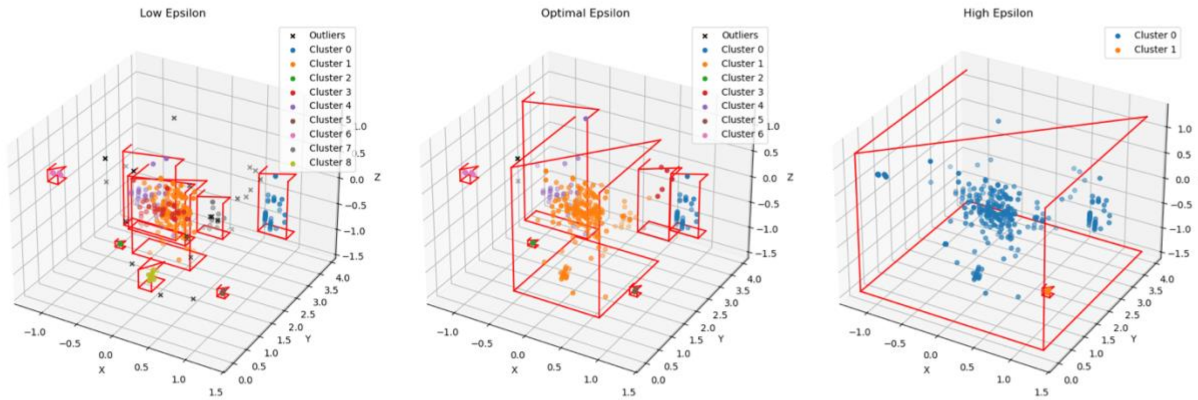


Figure 3 - Comparison of eps values with low, optimal and high values

3. Weight Analysis

Radar point clouds often represent human subjects as dense clusters surrounded by sparse noise due to environmental reflections and sensor limitations. To address this, (Zhao *et al.*, 2019) modified the traditional Euclidean distance metric by introducing a weighting parameter **α** to reduce the influence of vertical distances (z-axis), aligning better with human motion characteristics in the horizontal plane (x - y), this is denoted as;

$$D\left(p^{i}, p^{j}\right) = (p_x^i - p_x^j)^2 + (p_y^i - p_y^j)^2 + \alpha \,.\, (p_z^i - p_z^j)^2$$

Equation 4

To better understand the data's spatial distribution, it was first checked to confirm whether the points from the same cluster are more coherent in the horizontal (x-y) plane, as observed in real-world measurement studies. Specifically (Zhao *et al.*, 2019), found that human pose data points tend to be more clustered in the horizontal (x-y) plane but are more scattered along the vertical (z) axis, making it harder to merge them in the z-direction.

Figure_ below illustrates this spatial distribution across the coordinates. Which validates the finding of (Zhao *et al.*, 2019), where it was observed that the human pose data points are more clustered in the (x-y) plane and more scattered in the (x-z) plane.

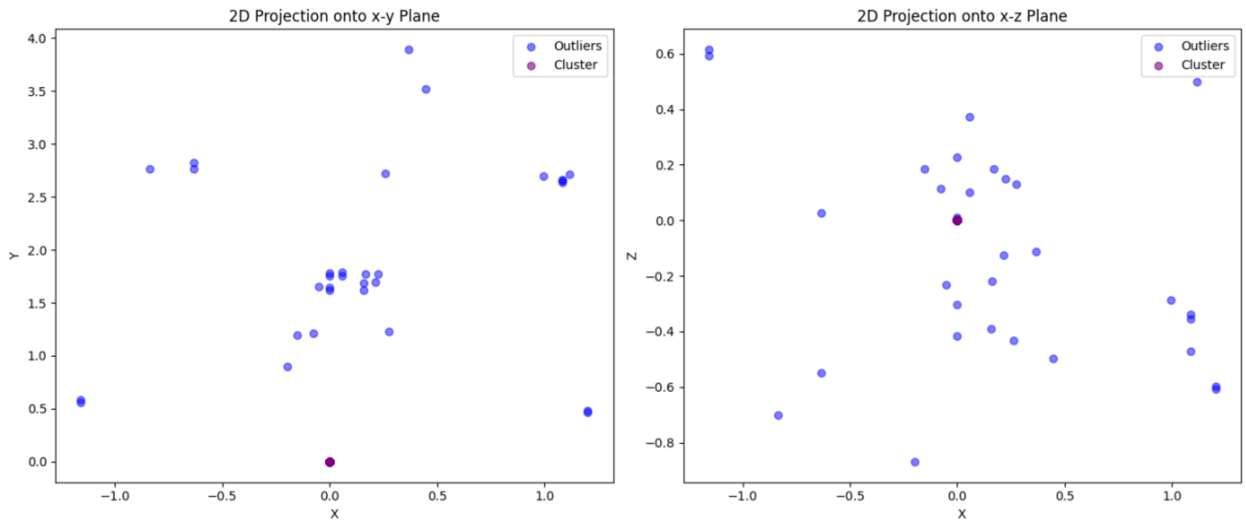


Figure 4 - Spatial Distribution of points across the x-y plane and x-z plane

To further analyse the implication of the weights, the impact of varying the weight values on the DBSCAN clustering results was observed. The bounding box highlights the yellow cluster, which represents the tightly grouped points which are all depicted in figure _ below.

- α = 0.0: The weight on z was set to 0.0, which begins to pull the points closer together in the Z-direction. The cluster's structure becomes more evident, and the bounding box fits more tightly around the yellow cluster. This adjustment improves the overall clarity of the cluster while still maintaining some spread.
- α = 0.5: The weight on Z is set to 0.5, which begins to pull the points closer together in the Z-direction. The cluster's structure becomes more evident, and the bounding box fits more tightly around the yellow cluster. This adjustment improves the overall clarity of the cluster while still maintaining some spread.
- α = 2.0: At this setting, the Z is heavily applied, causing the points to condense further in the Z-direction. The bounding box around the yellow cluster becomes the tightest, highlighting a more confined grouping. This shows the most compact and defined cluster, confirming that heavier Z-weighting enhances the grouping in the vertical direction.

These results reinforce our analysis, where adjust z helps to control the clustering's tightness and structure, particularly along the Z. The optimal αvalue depends on the data and the desired level of clustering coherence. Therefore 0.25 was selected as it is a middle ground.

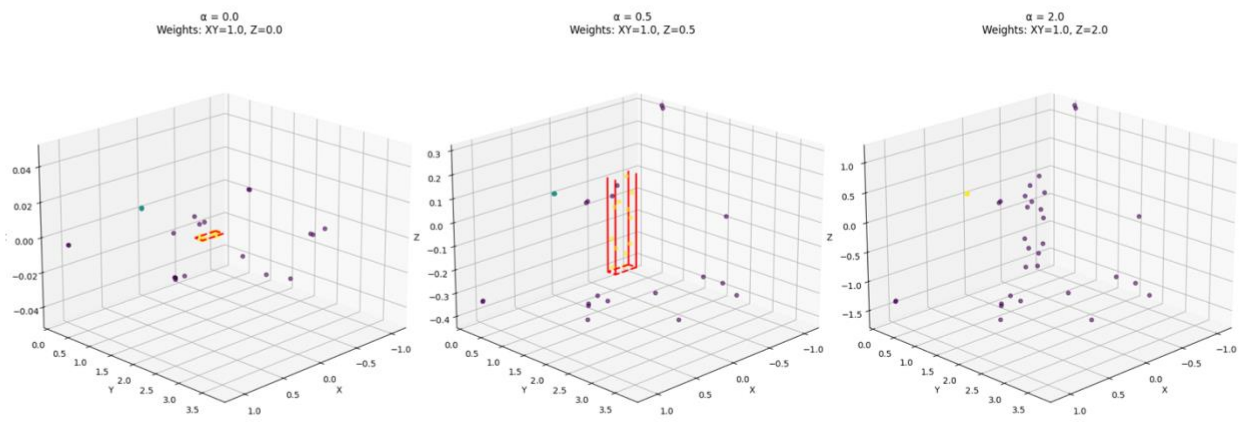


Figure 5 – Impacts of varying weights on the vertical axis

4. Optimal parameters

As shown in the analysis above, the optimal parameters for the DBSCAN are;

- Epsilon - 0.4
- Minsample - 6
- xyz weights - x,y =1.0, z = 0.25

Figure _ below shows the clustering using the optimal values

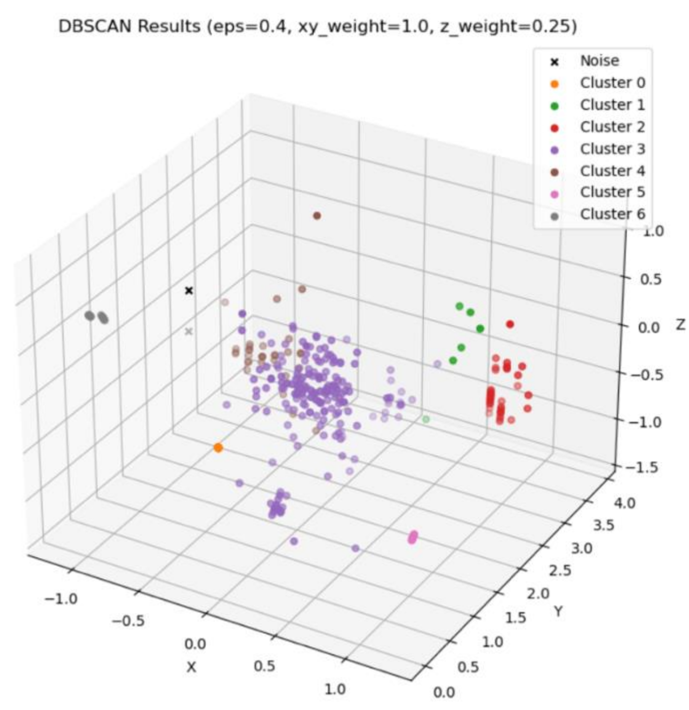


Figure 6 - DBSCAN result with optimal parameters

5.2.1 Hungarian Algorithm

i. Centroid Calculation

A centroid is the geometric centre of a set of points and for the purpose of this research, it refers to the location of a cluster points in 3D space. It is the arithmetic mean of all points in that cluster and can be represented as;

$$Centroid = \left(\frac{1}{n}\sum_{i=1}^{n} xi, \frac{1}{n}\sum_{i=1}^{n} yi, \frac{1}{n}\sum_{i=1}^{n} zi\right)$$

Equation 5 - Centroid

Where;

- Xi,yi, zi are the coordinated of the i-th point in the cluster
- n is the total nuber of points in the cluster

This was used to calculate the centroids of acquired clusters during implementation.

ii. Cost Matrix

Cost matrix is a 2D matrix used to represent the cost of associating clusters from different frames. It is done by computing the euclidean distance between centroids of clusters in different frames. This provides a way to quantify the similarity/distance between clusters across different frames. This was used for data association which is shown in the implementation

5.2.2 Kalman Filter

i. Parameter Selection

1. State and Observation vectors

a. State vector

The state was set to dim_x=6, tracking position (x, y, z) and velocity (vx, vy, vz). Position reflects to the physical location of detected actions, while velocity captures the motion dynamics. These parameters were chosen based on the dataset's characteristics, where actions involve low-intensity cardio exercises with minimal movement (low velocity) performed in a fixed area (position). This highlights the importance of accurately tracking position changes over time to capture subtle variations in motion.

b. Observation vector

The observation was set to dim_x= 3 which allows kalman to directly measure the position. This is allows the system to observe the position and estimate velocity using the state model, which is because there is minimal velocity change and can be inferred from positional changes over time which makes the position sufficient for tracking the motion dynamics.

2. Process and measurement noise optimisation

Further optimization involved tuning the process noise covariance (Q) and measurement noise covariance (R).

- Q: This parameter reflects the estimated accuracy of the system model itself. It accounts for the uncertainties in the prediction step of the filter, often related to the dynamics of the system being modelled.
- R: This parameter quantifies the expected noise inherent in the measurements obtained from sensors. It is crucial for weighting the trust in the sensor data relative to the model predictions.

These parameters are selected based on empirical evidence and iterative testing which resulted to the use of root mean square error (RMSE) to evaluate them where they were tuned. This resulted to 10.0 for Q and 0.05 for R.

Table 1 - RMSE for different combinations of process and measurement noise

Q	**R**	**RMSE**
0.01	0.05	0.2436
0.01	0.1	0.2703
0.01	0.2	0.2908
0.01	1.0	0.3210
0.1	0.05	0.1147
0.1	0.1	0.1572
0.1	0.2	0.1982
0.1	1.0	0.2705
1.0	0.05	0.0213
1.0	0.1	0.0386
1.0	0.2	0.0654
1.0	1.0	0.1573
10.0	**0.05**	**0.0024**
10.0	0.1	0.0047
10.0	0.2	0.0091
10.0	1.0	0.0386

5.2.2 Data Processing Techniques

i. Noise removal

To filter out irrelevant data points, mainly points that are close to origin, a threshold distance was applied from the origin using the **L2 norm** to remove them. This was done because the initial data processing padded the data with zeros, resulting in the accumulation of points at zero origin. This removed irrelevant points.

ii. Segmentation

As the aim was to do interframe tracking with the hungarian, the frame was split into different segments to check the most effective. With 1100 points the points can be divided into 2, 4, and 10 as the frames were split into those variables. This allows for a more focused analysis of human clusters. Their states were analysed to select the best option for implementation

- 2 segments
- 4 segments
- 10 segments

5.5 Implementation (How I implemented them)

This section details how the research was implemented through the flowchart that was shown earlier.

5.2.1 Data loading and preprocessing

To initiate the implementation, the earlier processed data was loaded using picke module. Also being that the dataset of interest was new_x from the training set, the first 3D of the radar points were extracted to create a consistent dataset.

Next, the data was split into temporal segments for further analysis. Specifically, the dataset was divided into 2 segments, as detailed in the methodology. This is shown in the figure – below.

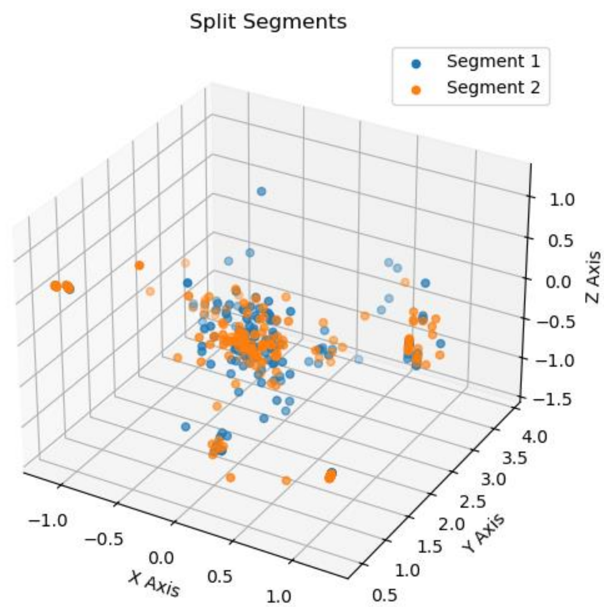


Figure 7 - Data points after frame splitting in segments

5.2.2 Clustering

DBSCAN was then applied on each segment independently to identify clusters within each frame. This was done using the optimal parameters. And this yielded into clusters as seen in the picture below.

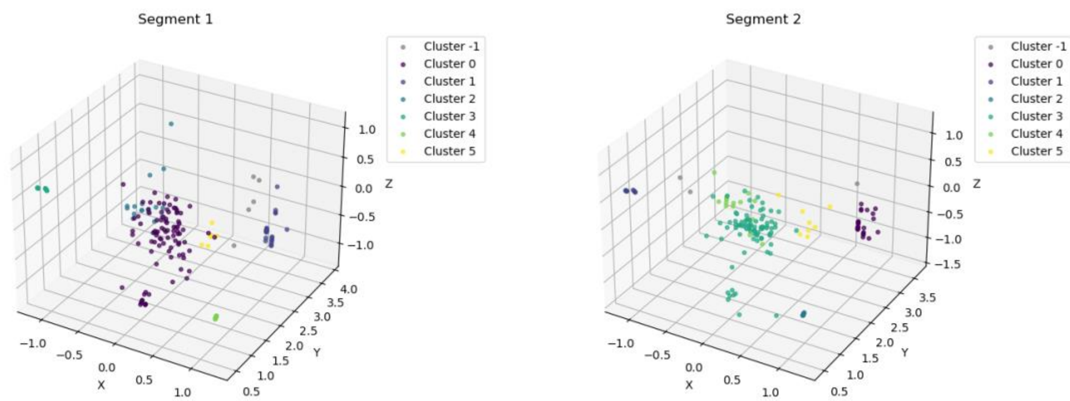


Figure 8 - DBSCAN on 2 segments after splitting

5.2.3 Track Assignment

After clustering, the centroids were calculated and then used to create a cost matrix using the **scipy.spatial.distance.cdist** function. as seen below. The cost matrix involbed the application of the Hungarian algorithim by the use of use **linear_sum_assignment** from scipy.optimize to solve the assignment problem and create the cost matrix, which gives an output of row_indices, col_indices, the optimal assignments of clusters across frames, minimizing the total distance. And they were then identified as tracks.

This matrix is seen below with the matches identified. The matches made between clusters is also shown below

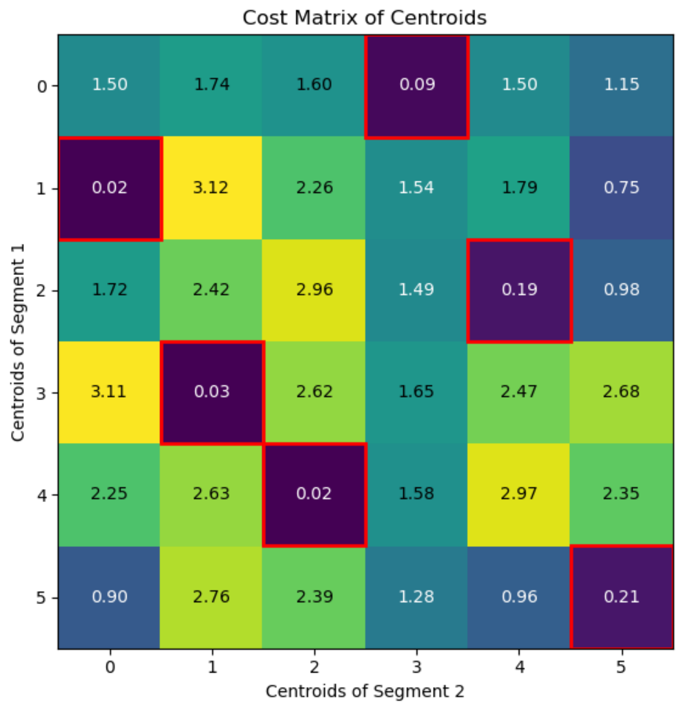


Figure 9 - Cost Matrix

5.2.4 Kalman Filtering

To track the motion of the clusters across segments, the kalman was initialised and applied. Ensuring that the model could predict the future state based on the previous observations. The Kalman filter was updated with observations (centroids of the clusters), refining the predictions over time.

5.2.5 Human Action Recognition

Once the tracks were assigned and updated using the kalman, the next step was to evaluate and select the best match for the human actions based on RMSE, Velocity Change, and Duration all derived from the Kalman.

- RMSE: this measures the difference between the predicted and observed positions of the tracked objects and was chosen to assess the accuracy of the Kalman's predictions. As the system involves slight movements, it is crucial to measure how closely the predicted positions match the observed positions across the frame. A lower RMSE indicated the predictions are more aligned to the data. The formula used to compute this was;

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(predicted\ position_i - actual\ position_j)^2}$$

Where:

- N is the number of observations
- Predicted position is the output from the Kalman filter for each frame.
- Actual position is the observed position from the dataset.

- Velocity Change: This measures the difference in the tracked objects' velocity over time. Large, sudden changes indicated noise or errors in the track as the dataset has minimal movement. It was calculated using;

$$Velocity\ change = \frac{1}{N-1}\sum_{i=1}^{N-1}|velocity_i - velocity_{i+1}|$$

Where:

- velocity_i is the velocity of the track at frame i.
- N is the total number of frames in the track.

- Duration: this indicated how long the track lasted, longer tracks with stable and consistent motion indicate more reliable tracking. This was computed as;

$$Duration\ score = duration\ of\ the\ track * 0.1$$

The constant 0.10.1 was chosen to normalize the duration score and balance it with the other parameters (RMSE and velocity change).

To determine the best match for the human action, a weighted score was calculated for each track using the three parameters. The weighted formula is designed to give more importance to position accuracy (RMSE) and velocity consistency (velocity change), with a smaller contribution from duration.

The final score for each track is computed as:

$$Track\ Score = \left(\omega_{velocity} * \frac{1}{1 + velocity\ change}\right) + \left(\omega_{rmse} * \frac{1}{1 + rmse}\right) + (\omega_{duration} * duration)$$

Where:

- $\omega_{velocity}$, ω_{rmse}, and $\omega_{duration}$ are the weights assigned to the respective parameters, and they were selected to balance the contributions of each parameter based on their relative importance in determining the best track match.

To ensure robustness, different weight combinations were tried and were evaluated on the final track selection and overall performance of the model. By varying these weights, I aimed to explore how sensitive the model is to changes in the importance of each parameter and identify the most effective combination for accurately selecting the best matching track. The evaluation and comparison of these different weight configurations are discussed in detail in the evaluation chapter.

5.2.6 Evaluation and cleaning

After identifying the best matches, the final step involved evaluation and cleaning of the tracks. The tracks were evaluated based on the calculated scores, and any tracks that did not meet a minimum threshold were padded as zero ensuring a clean output of recognized action.

5.6 Technical Challenges

Evaluation

- Different weights
- Show before dbscan, after, and the best one, the 4 plot
- Evaluate accuracies after etc
- Show the state and results of Kalman
- Show the best human action and its characteristics
- Limitations
- Future work
- Ethical considerations