

Beam Prediction for mmWave V2I Communication aided by Geolocation and Machine Learning

Sherif Adeshina Busari^{1,*}, Ifiok Otung¹, Muhammad Ali¹ and Raed Abd-Alhameed^{1,2}

University of Bradford, UK

s.a.busari@bradford.ac.uk; i.otung@bradford.ac.uk; m.ali70@bradford.ac.uk; r.a.a.abd@bradford.ac.uk

Al-Farqadein University College, Iraq

r.a.a.abd@bradford.ac.uk

*Correspondence: s.a.busari@bradford.ac.uk

Received: 21 July 2025; Accepted: 6 August 2025; Published: 25 October 2025

Abstract: Millimetre wave (mmWave) systems require high beamforming gains to overcome the unfavourable impacts of high path losses at mmWave frequencies. Large antenna arrays enable such gains through highly directive narrow beams which then require multiple beams to cover the spatial directions of interest. The required beam management for such systems, particularly for mobile use cases such as the vehicle-to-infrastructure (V2I) scenarios, is challenging. Real-time optimal beam selection from codebooks consumes radio resources and incurs large training overheads. As a result, geolocation side information and machine learning (ML) algorithms are being explored to address beam management challenges. However, prior works have mostly applied their solutions using simulations that are based on synthetic datasets. Recently, real-world datasets based on extensive mmWave measurements have become available. Leveraging the real-world datasets, in this work, we evaluate and compare the performance of three ML (i.e., k-nearest neighbours, support vector machine and decision tree) algorithms on mmWave V2I beam selection aided by global positioning system latitude and longitude coordinates as the only two features for the ML. The results show the impact of codebook sizes on the accuracies of the ML algorithms under ten different scenarios. The results also reveal the limitations of the geolocation-aided beam prediction as average accuracy could go below 30% in some scenarios, and higher than 90% in other scenarios. These performance results point to the need for multi-modal approaches (involving a combination of different sensors' data) for efficient mmWave V2I beam prediction.

Keywords: Beam Prediction; Decision Tree; GPS; K-Nearest Neighbour; Machine learning; mmWave; Support Vector Machine; V2I

1. Introduction

Millimetre-wave (mmWave) systems enable enhanced mobile broadband communication by exploiting larger bandwidth than available at the legacy sub-7 GHz frequency bands. However, to facilitate the desired multi-Gigabits-per-second (Gbps) rates, mmWave systems need to combat the high path losses (PL) at such high frequencies [1]. For example, according to the Friis equation, the free space path loss (FSPL) for a 6 GHz system at a transmitter (TX)-to-receiver (RX) separation distance of 100 m is 88 dB. This same 88 dB is the FSPL at only 10 m separation distance for 60 GHz mmWave propagation. Therefore, to extend range and facilitate reliable communication, mmWave systems employ large antenna arrays with narrow beams that offer high beamforming gains and directivity to ensure sufficient received signal power. This solution is, however, not without its own challenges with respect to beam management, particularly for vehicle-to-everything (V2X) scenarios with mobility of both the TX and RX or either one of them [2-4].

The challenge with narrow beams is that they cover limited spatial directions. This, therefore, necessitate the use of either multiple beams with predefined beamforming codebook [5-6] to cover the entire region of interest or field of view (FOV), or beam sweeping operations where beams cover a spatial area during a time instance in a predetermined way, and sweeps through another area in another time instance [3], using approaches based on the beam's angle of arrival (AoA) and angle of departure (AoD) [7]. In addition, the beam management procedures (i.e., beam alignment, tracking, training, selection and steering) typically consume radio resources and are associated with large training overheads [2, 6].

To address the radio resource challenge, many authors consider alternatives such as using geolocation side information (i.e., localisation and positioning systems such as the global positioning system (GPS)) in aiding beam management procedures and in reducing beam training overhead [2-3, 8]. In addition to these system-aiding alternatives, many authors also explore artificial intelligence (AI)-based solutions such as machine learning (ML) or deep learning (DL) in tackling beam management challenges [2]. Therefore, while classical approaches have been employed over the years, there is a growing adoption of the "AI/ML for wireless" paradigm in communication systems where ML techniques are used to tackle several challenges in communications systems, including beam management [9].

In [8], the authors employed ray tracing-based simulations to investigate the performance of support vector machine (SVM)-aided beam management for 5G new radio mmWave systems. The simulations used geolocation side information to reduce the required channel state information (CSI) feedback leveraging the proposed scheduler and using sum rate, latency and overhead as metrics. A similar ray-traced simulation was undertaken in [10] that leveraged GPS signals for beam alignment in 28 GHz vehicular network setups. The study employed random forest classifier and multilayer perceptron as two supervised classification ML algorithms, compared against the baseline naïve context information (CI) algorithm. Performance was evaluated using accuracy, precision and recall for the optimal beam prediction challenge. The authors of [11] and [12] went further by employing DL for beam prediction, alongside the ML-based approaches. A common denominator in these state-of-the-art works is that the ML, DL and GPS-aided frameworks are applied on synthetic datasets which present some limitations when compared to real-world experimental data.

Different from earlier approaches, the authors of [6] proposed position-aided beam prediction frameworks that use GPS coordinates for beam selection at the infrastructure in vehicle-to-infrastructure (V2I) scenarios. The authors considered three approaches for performance evaluation: (i) look-up table, (ii) k-nearest neighbours (KNN), and (iii) fully connected neural network. Interestingly, the frameworks are tested using large real-world datasets¹ curated from extensive 60 GHz mmWave experiments under different scenarios and use cases (i.e., DeepSense6G) [13]. In [6], the results across nine different scenarios (Scenarios 1-9) show average beam prediction accuracy less than 40% for a 64-beam codebook and 80% for the downsampled 8-beam codebook. These results are less than the typical over 95% accuracy from synthetic datasets and this underscores how the proposed ML-based and GPS-aided solutions perform on real-world datasets. The solutions in [6] leave room to test other ML algorithms on the considered datasets and to consider other scenarios to draw further performance insights.

The authors of [14] investigated the impacts of dataset and codebook sizes on V2I beam prediction using four different algorithms (KNN, SVM, DT and Naïve Bayes (NB)), using also the experimental datasets from DeepSense6G¹. The authors investigated the impacts of the ML training-to-testing split ratios (80:20, 70:30 and 60:40%) and beam codebook sizes ($Q = 8, 16, 32$ and 64). The authors of [15] similarly employed KNN, SVM, DT and NB algorithms for five scenarios (i.e., Scenarios 1, 2, 5-7 of the DeepSense6G dataset), and further employed other metrics such as confusion matrices, area under the receiver operating characteristic curves, precision, recall, specificity and F1-score for performance evaluation. The work in [15] also investigated the impacts of data splits and codebook sizes on the beam prediction accuracy. However, unlike this work, the works in [14-15] only considered five scenarios (1, 2, 5-7) and did not also consider the day and night characteristics of the datasets. The overall number of dataset samples in [14-15] are also considerably lower than the number of dataset samples used in this work.

¹DeepSense6G, 'A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset for 6G Deep Learning Research'. [Online]. Available: <https://www.deepsense6g.net/>

In this work, therefore, we compare the performance of three algorithms on ten V2I scenarios from the DeepSense6G datasets¹. The ML algorithms are the KNN, SVM and the Decision Tree (DT). The ten scenarios (Scenarios 1-2, 5-7, 14, 31-34) represent different data collection locations and time of the day and have different datapoints or samples, as further described in subsection 2.2 (and presented in Table 2 and Table 2). A sample scenario (i.e., Scenario 7) is shown in Figure 1 and described in subsection 2.1.

To the best knowledge of the authors, the KNN has been explored in [6,14,15] and the SVM and DT algorithms have been considered in [14-15] for position-aided ML-driven beam prediction challenge using data from the real-world DeepSense6G datasets. However, the considered numbers of scenarios, and by extension the number of data samples in [6, 14-15] are limited. In addition, this study investigates the impacts of several parameters such as beam codebook size, number of data samples, time of the day aggregates, etc., on system performance considering the KNN, SVM and DT algorithms and thus provides further insights of the performance of the algorithms.

The remainder of this paper is organised as follows. In Section 2, we present the system model that describes the network deployment layout, datasets and preprocessing operations. Section 3 then presents the considered ML algorithms. In Section 4, we present the results and discussion while the conclusions and future research directions are presented in Section 5.

2. System Model

In this section, we describe the considered deployment layouts, datasets and preprocessing operations on the datasets. Figure 1 shows a sample of the network deployment layout (i.e., Scenario 7). Other scenarios show similar network layouts but at different locations, and with different number of lanes and data samples.



Figure 1. GPS View of Network Deployment (Scenario 7)¹

2.1. Deployment Layout

As shown in Figure 1, we consider a V2I system where the TX is a vehicle equipped with a single ($N_t = 1$) omnidirectional antenna at 60 GHz, and a GPS sensor. Scenario 1 features two lanes where the vehicle (TX) traverses the lanes multiple times in both directions and sends both communication and GPS signals to the street-level base station (BS) or access point (AP) infrastructure on the sidewalk. The infrastructure then predicts the beam with the highest received power from its beam codebook for each sample point. Other scenarios have similar V2I deployment setups as described on the DeepSense6G dataset repository¹.

The infrastructure is equipped with a phased array employing analog beamforming. The BS (i.e., RX) features $N_r = 16$ uniform linear array (ULA) 60 GHz antennas and a GPS sensor receiver. The BS array employs an oversampled codebook with $Q = 64$ receive beams or beamforming vectors $\mathbf{w}_q \in \mathbb{C}^{N_r \times 1}$ or $\mathbf{W} \in \mathbb{C}^{N_r \times Q}$ such that $\mathcal{W} = \{\mathbf{w}_q\}_{q=1}^Q$. The transmit signal $x \in \mathbb{C}^{1 \times 1}$ and received signal vector $\mathbf{y} \in \mathbb{C}^{Q \times 1}$ are related by (1):

$$\mathbf{y}_q = \mathbf{w}_q^T \mathbf{h}_q^T f x + \mathbf{w}_q^T \mathbf{n} \quad (1)$$

where $\mathbf{h}_q \in \mathbb{C}^{N_r \times 1}$ is the complex channel vector that holds the amplitude and phase transformations that occur between each BS antenna and the UE antenna, f is the TX beamformer and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ represents a complex normally distributed noise. The beamforming gain is $G = \mathbf{a}_{ULA}(\varphi_q)^T \mathbf{h}$ where the array response vector $\mathbf{a}_{ULA}$ is given by (2):

$$\mathbf{a}_{ULA}(\varphi_q) = \frac{1}{\sqrt{N_r}} [1, e^{jld \sin \varphi_q}, \dots, e^{j(N-1)ld \sin \varphi_q}] \quad (2)$$

where beam q is from the codebook Q , and $\varphi \in [-\pi/3, \pi/3]$ represent the field of view (FoV) or φ_q represents the beamwidth of beam q , λ is the wavelength and d is the inter-element spacing of the AP's ULA and $l = 2\pi/\lambda$.

The optimal beam selection problem corresponds to the selection of the beamforming vector that achieves the highest receive power at the infrastructure. This is formulated as in (3). However, since the CSI acquisition in highly mobile scenarios is challenging, real-time position coordinates can be used to aid the beam prediction challenge. This is what we explore in this study. Further details on the DeepSense6G experimental testbed are available in [13] and on the DeepSense6G dataset repository¹.

$$\mathbf{w}^* = \arg \max_{\mathbf{w} \in \mathcal{W}} |\mathbf{w}^T \mathbf{h}^T f|^2 \quad (3)$$

2.2. Scenarios' Datasets and Preprocessing

The dataset used for this study are from the DeepSense6G multimodal open datasets¹ for mmWave communication [13]. The dataset is a large multimodal dataset with several scenarios, data samples and measurement units (i.e., devices) and measured variables. The dataset is well documented on its website such that users can easily access the needed information for their specific use case or scenario. The full dataset contains many measured variables relating to data index, GPS values, mmWave power, beam index, LiDAR, radar, camera images, number of satellites used, etc. The dataset contains the following headers, which are adequately described on the DeepSense6G repository¹:

`['index', 'unit1_rgb', 'unit1_pwr_60ghz', 'unit1_loc', 'unit1_lidar', 'unit1_lidar_SCR',
'unit1_radar', 'unit1_beam_index', 'seq_index', 'unit2_loc', 'unit2_direction',
'time_stamp', 'unit2_sat_used', 'unit2_fix_type', 'unit2_DGPS', 'unit2_PDOP', 'unit2_HDOP']`

However, in this work, we have only fetched the specific variables/columns in Table 1 below for the use in this work, where unit1 is the AP/infrastructure and unit2 is the vehicle (TX). The "index" refers to the dataset sample number in each scenario, each row of the "unit_pwr_60ghz" column contains the list of measured mmWave power in the 64 beams, each row of the "unit1_loc" contains the infrastructure's GPS coordinates (which is same for each scenario as the infrastructure (unit1) is static for each scenario) while each row of "unit2_loc" column contains the GPS coordinates for the vehicle which is different per row as the vehicle (TX) is mobile. The beam with the highest power among the 64 beams is the ground truth index or class. The measured GPS coordinates for both unit1 (infrastructure or RX) and unit2 (vehicle or TX) allows conversion to cartesian coordinates as well as the calculation of the Euclidean distance between corresponding TX and RX positions for each row or datapoint as used in KNN for example or for the determining the hyperplane as employed in SVM.

Table 1. Employed dataset features

index	unit1_pwr_60ghz	unit1_loc	unit2_loc
1	./unit1/mmWave_data/mmWave_power_0.txt	./unit1/GPS_data/gps_location.txt	./unit2/GPS_data/gps_location_0.txt
2	./unit1/mmWave_data/mmWave_power_1.txt	./unit1/GPS_data/gps_location.txt	./unit2/GPS_data/gps_location_1.txt
3	./unit1/mmWave_data/mmWave_power_2.txt	./unit1/GPS_data/gps_location.txt	./unit2/GPS_data/gps_location_2.txt
4	./unit1/mmWave_data/mmWave_power_3.txt	./unit1/GPS_data/gps_location.txt	./unit2/GPS_data/gps_location_3.txt
.	.	.	.
.	.	.	.
.	.	.	.

On data preprocessing, we convert the GPS latitude and longitude tuples into Cartesian coordinate values. The two-dimensional (2D) Cartesian location values are used as the two features for the ML-based beam prediction or selection algorithms. This feature extraction stage is equivalent to having a GPS-to-Cartesian coordinates converter before the ML prediction module. Also, the Cartesian location coordinates are normalised before use as ML features, following standard ML practice.

The dataset currently consists of forty-four different scenarios that cover different use cases. Out of these, V2I use cases are covered by 17 scenarios (i.e., Scenarios 1–9, 13–15, and 31–35). Scenarios 10–12 are for pedestrian communications, Scenario 16 is for indoor communications, Scenario 23 is for drone communications, Scenarios 36–39 are for vehicle-to-vehicle (V2V) communication, and Scenarios 42–44 are for integrated sensing and communication (ISAC) use cases. Throughout this manuscript, we have retained the scenario numbering exactly as used in the DeepSense6G dataset¹. This is to ensure consistency with the source of the datasets and to ease performance comparison with other works that employ the same datasets.

Considering the V2I scenarios, we have focused on only ten scenarios (i.e., Scenarios 1-2, 5-7, 14, 31-34). These selected scenarios are without any missing data, thus easing the preprocessing and performance evaluation. We have also considered only the mmWave communication and GPS data (i.e., without other sensor data such as Light Detection and Ranging (LiDAR), Radio Detection and Ranging (RADAR), and camera). Typically, vehicle localisation employs GPS signals when available and resorts to sensed data (e.g., LiDAR, RADAR, and camera images, etc) when GPS signals are not available [16]. In addition, the approach to employ only GPS signals in this work is to limit the required computational complexity of the system, as processing the position-based features is less computationally demanding than using camera/vision-based beam prediction that requires image processing, for example, or using the multi-modal approach that involves a combination of sensed data, which improves the system's accuracy but at the expense of higher computational demand.

A summary on each of the considered ten scenarios, together with the total data (i.e., with all ten scenarios' samples combined) is given in Table 2. Also, the period of the day and weather condition under which the measurements were taken are also presented in Table 2. Also, each scenario dataset is split into 80% training and 20% test samples using the hold-out partitioning method. The partitioning is randomised in each iteration, and the results are averaged over ten iterations, considered sufficient for the beam prediction use case under consideration.

Table 2. Summary of scenarios' datasets

Scenario #*	Time of the Day	Weather Condition	Number of Samples		
			Total	Training (80%)	Testing (20%)
1	Day	Clear	2411	1929	482
2	Night	Clear	2974	2380	594
5	Night	Rainy	2300	1840	460
6	Day	Clear	915	732	183
7	Day	Clear	854	684	170
14	Night	Clear	512	409	103
31	Day	Clear	7012	5609	1403
32	Day	Clear	3235	2588	647
33	Night	Clear	3981	3184	797
34	Night	Clear	4439	3551	888
Total	Mixed	Mixed	28633	22906	5727

*We have retained the scenario numbering as in DeepSense6G¹.

2.3. Beam Codebook Downsampling

The original codebook considered in this study is an oversampled codebook with $Q = 64$ beams. With the AP's FoV, we show in Figure 2² and Figure 3³, for the ideal and real beam patterns respectively, that these 64-beam codebook has significant overlap. It is also more computationally complex due to the high

²The ideal beam pattern is generated using the MIMO4MATLAB Toolbox available at: <http://mimoformatlab.com>

³The real beam pattern is generated using the DeepSense6G measurement data and code available at: <https://www.deepsense6g.net/data-collection/> and <https://www.deepsense6g.net/tutorials/>, respectively.

dimensional search space. For these reasons, we downsample the codebook from $Q = 64$ beams to $Q \in \{32, 16, 8\}$ beams. In addition, a comparison of Figure 2 and Figure 3 underscores the difference between synthetic and real-world datasets.

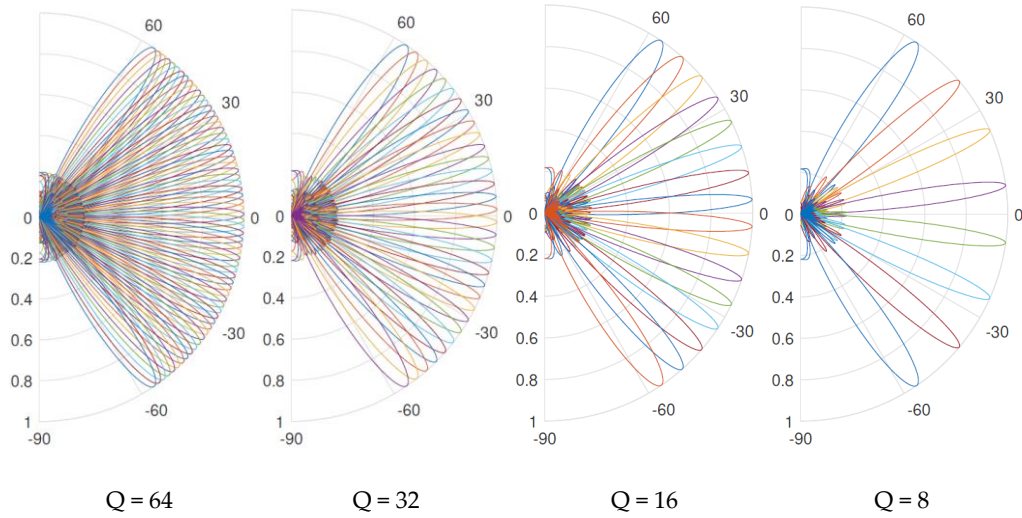


Figure 2. Ideal multi-beam patterns for $Q \in \{64, 32, 16, 8\}$ [15]

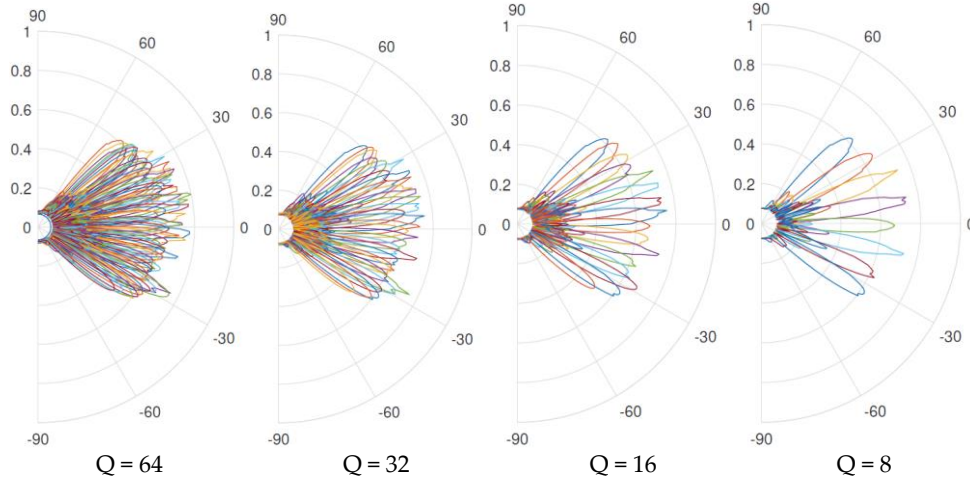


Figure 3. Real multi-beam patterns for $Q \in \{64, 32, 16, 8\}$ [15]

3. ML Algorithms

In this work, we consider three supervised learning classification algorithms for mmWave V2I beam prediction. The three algorithms are described as follows:

3.1. k-Nearest Neighbours (KNN)

KNN is a simple, easy-to-implement and widely-used supervised learning classification algorithm where a sample is classified based on a specific number of its nearest neighbours [14]. It is a lazy learner that works by learning the training dataset and thereafter determining the label of the new sample based on the labels of its closest neighbours. KNN considers that the nearby samples should have the same label. There are several methods employed in determining the k value (i.e., the number of neighbours). The common method is, however, by trial and error where small, odd numbers are tested one after the other to obtain desirable results. Also, several distance metrics (such as the Euclidean, Minkowski, Hamming and Manhattan distances, etc) are employed to compute the sample's closest/nearest neighbours [17]. With a predefined beam codebook, each beam covers a spatial direction within the FoV. Thus, KNN considers that similar or neighbour positions (using the location coordinate tuples) should have similar beams. Following the implementation in [6], the mode of the beams from N_{knn} nearest neighbours is selected as the predicted beam. The k -smallest difference in Euclidean distances is employed as the

parameter used in selecting the neighbours and predicting the beams. The true beam (ground truth) is the beam with the highest measured received power for each sample TX point.

3.2. Support Vector Machine (SVM)

SVMs are efficient and powerful classifiers that have found application in diverse real-world use cases due to their versatility and extraordinary generalization capability. SVM can be adapted for the specific use case leading to variants such as binary SVM and multi-class SVM, or for the type of dataset under consideration leading to variants such as SVM for unbalanced datasets, SVM for large datasets, etc [18]. The SVM algorithm employed in this work is a multi-class SVM based on the directed acyclic graph (DAG). The DAG-SVM was proposed in [19] and we adopt the MATLAB implementation code⁴. The DAG-SVM predicts beams using only two features (location coordinates tuples in this case) and employs the one-versus-one framework to transform many two-class classifiers into a multi-class classifier for the multi-beam/multi-label prediction challenge. For a Q -class problem, the DAG-SVM contains $Q(Q-1)/2$ classifiers, one for each pair of classes. It employs a Gaussian radial basis function kernel to generate nonlinear boundaries or SVM hyperplane between the two classes of each combination [20]. The DAG-SVM is hereinafter simply referred to as SVM. Despite its optimal solution and high discriminative power, a challenge with SVM is that it generates a large amount of data that can become prohibitive, particularly when the number of classes is high [18].

3.3. Decision Tree (DT)

The Decision Tree (DT) algorithm is a rule-based classifier that employs a top-down classification structure. In its tree structure, each feature is represented by a tree vertex, and each tree branch shows the value of the feature. The tree's topmost vertex is known as the root of the decision tree while the vertices at the bottom are known as leaves, with each leaf representing a class. The DT employs information gain or entropy differences for classification [20]. For the beam prediction application in this work, the DT supervised learning algorithm predicts the target beam by learning simple decision rules that it infers from the data features. The tree works as a piecewise constant approximation, can handle multi-class output problems and can efficiently deal with large and complicated datasets, while making no assumption of the underlying data distribution [14,21]. The DT classifier predicts the class of sample with the highest probability. In the event of multiple classes with same and highest probabilities, the class with the lowest index among the classes is predicted. The complexity of the tree determines its accuracy.

4. Results and Discussion

In this section, we evaluate the performance of the ML algorithms on the scenarios' dataset. We employ beam prediction accuracy as the performance metric. It should be noted that throughout this paper, the beam prediction accuracy refers to the testing accuracy. It is the percentage of test samples where the predicted beam by the ML algorithms (KNN, SVM and DT) are the same as the ground truth/measured optimal beam at the AP.

For KNN, we also evaluate the top- k accuracy, with $k \in \{1, 2, 3, 4, 5\}$ and which measures the ratio of the test samples where the ground truth/measured optimal beam is within the top- k predicted beams. It is the percentage of hitting the optimal beam by searching the top- k beams.

4.1. Performance Comparison of Algorithms

Figure 4 shows the beam prediction accuracies of KNN, SVM and DT for downsampled codebooks $Q = 32, 16$ and 8 for the ten considered scenarios. The trends in Figure 4 show that the lower the beam codebook size Q , the better the beam prediction accuracy. This means that lower number of beams in the codebook with wider beamwidth per beam is beneficial for more accurate beam pairing predictions. This is, however, at the expense of the higher array gain realisable with narrower beams. The difference between the ideal beam patterns and the real beam patterns as illustrated in Figure 2 and Figure 3,

⁴<https://www.mathworks.com/matlabcentral/fileexchange/65232-binary-and-multi-class-svm>

respectively, underscores the difference between synthetic and real-world datasets. Also, the low performance results are majorly for high codebook values (e.g., 32 beams and 64 beams) where due to significant overlap of the beams, the probability of misclassifying the optimal beams becomes higher, where the probability that the algorithm predicts near-optimal beams is higher, as against predicting the optimal beams or ground truth index which then leads to higher misclassification by the algorithms as the codebook size increases.

Also, for all considered scenarios, the results also show that SVM marginally outperforms both KNN and DT. However, for all scenarios, codebook sizes and algorithms, no correlation can be established between the number of samples in each scenario and the accuracy. For example, as shown in Table 2, Scenario 14 has the lowest number of datapoints whereas Scenario 7 has the least performance while not having the highest number of data samples. Along the same line, Scenario 31 with highest number of data samples does not show any performance superiority over most of the scenarios with lower data samples. Overall, the results show that the beam prediction accuracy is scenario-specific. However, the results show that the accuracy is codebook size-dependent, as the higher the codebook size, the lower the prediction accuracy.

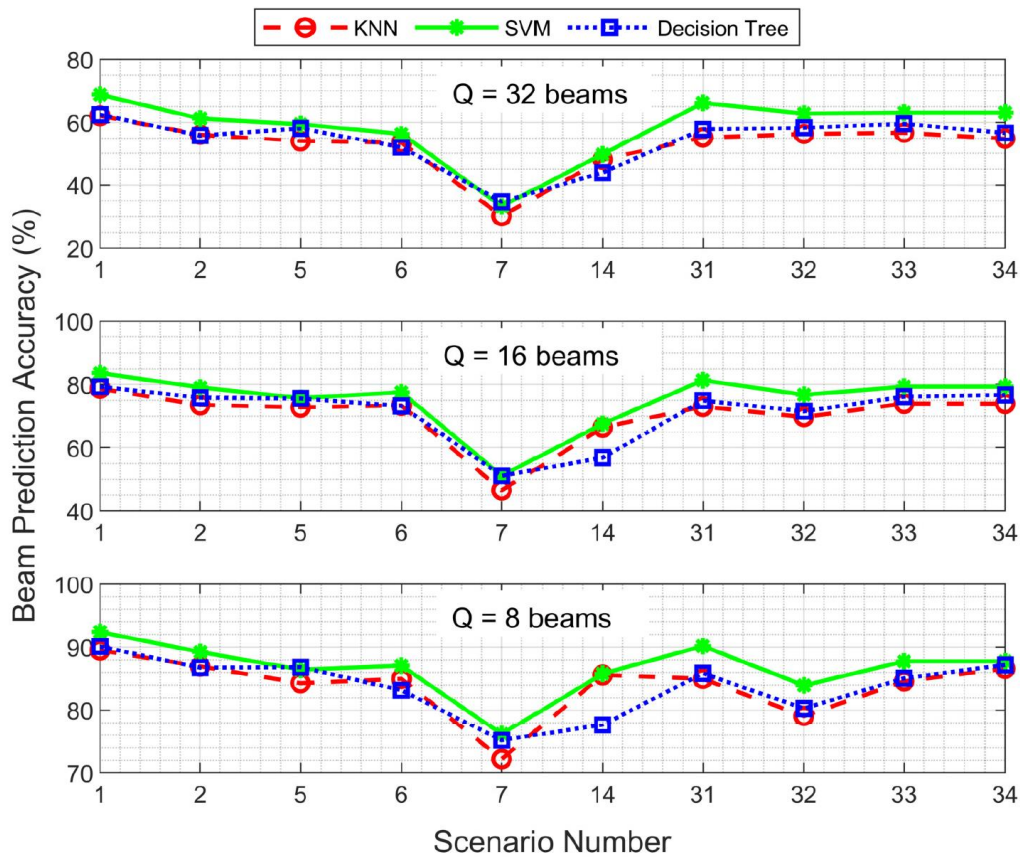
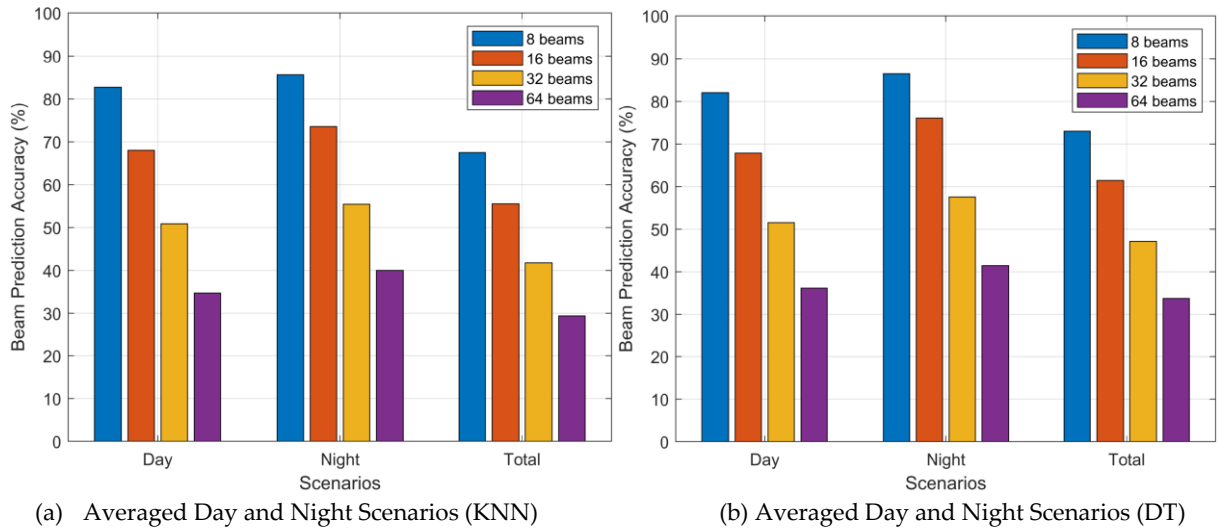


Figure 4. Algorithms' Accuracies for $Q = 32, 16$ and 8

4.2. Performance with Day and Night Aggregates

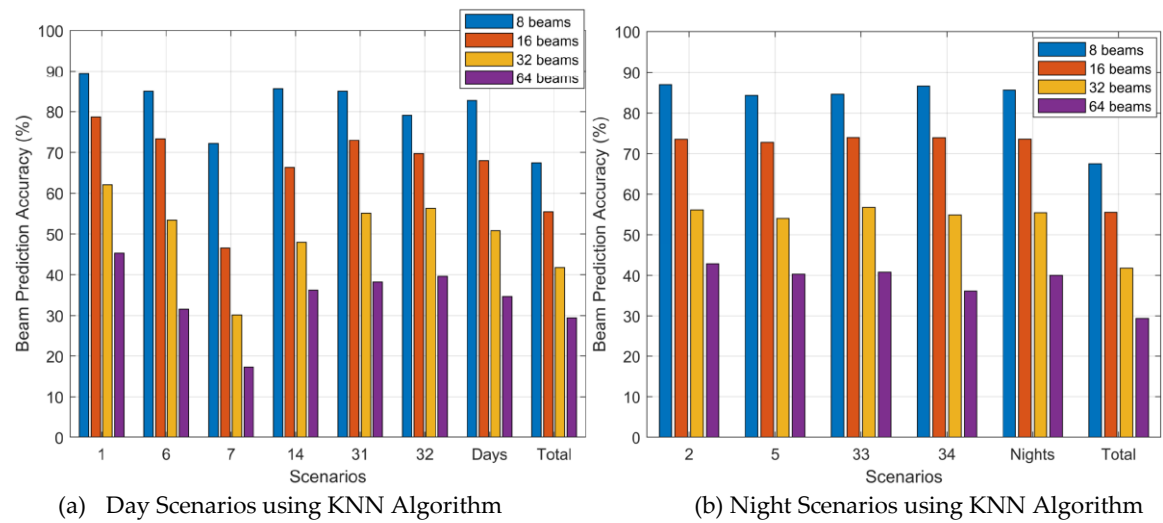
In Figure 5(a) (KNN) and Figure 5(b) (DT), we compare the beam prediction accuracy for aggregated day scenarios (#1, 6, 7, 14, 31 and 32) and aggregated night scenarios (#2, 5, 33 and 34), against the total that includes all scenarios (i.e., containing all day and night scenarios).

Following the same trend as in Figure 4, the results in Figure 5 also show that the lower the Q , the higher the accuracy. Also, the night scenarios show higher prediction accuracies than the day scenarios, while the total scenarios show the least accuracies. No correlation could be established from the day and night statistics, except that the night performance statistics were averaged over 4 scenarios, day scenarios were averaged over 6 scenarios while the total scenarios were averaged over 10 scenarios, where the number of the respective aggregates (day, night and total) could have influence on the performance statistics.



(a) Averaged Day and Night Scenarios (KNN)

(b) Averaged Day and Night Scenarios (DT)

Figure 5. Averaged Scenario Performance for KNN and DT for different codebook sizes and Time of the Day.

(a) Day Scenarios using KNN Algorithm

(b) Night Scenarios using KNN Algorithm

Figure 6. Day and Night Scenarios' Performance for different codebook sizes using KNN

Figure 6(a) shows day scenario-per-scenario, all “days” scenarios combined and all scenario combined (i.e., total) performances using KNN and for $Q = 8$ to $Q = 64$. Similarly, Figure 6(b) shows night scenario-per-scenario, all “night” scenarios combined and all scenario combined (i.e., total) performances using KNN and for $Q = 8$ to $Q = 64$. The results follow the same trends as those of Figure 5.

4.3. Top-k Accuracy using KNN

Figure 7 shows the top-k accuracy performance using KNN across the ten scenarios considered for $Q = 8, 16, 32$ and 64 . The results show top-1 accuracies of 20-42% for $Q = 64$ and 70-90% for $Q = 8$. For the top-5 accuracy, the results in Figure 7 show accuracies of 60-82% and 95-98% for $Q = 64$ and $Q = 8$, respectively. The results show that the higher the number of “ k ” neighbours, the higher the accuracy for the considered scenarios and values of k . A tradeoff between Q and k is therefore important for desired system performance. Overall, these results are consistent with the outcomes in [6, 14-15] for Scenarios 1-9. However, this work also provides results for Scenarios 14 and 31-34 that were not covered in [6, 14-15].

5. Conclusion

In this paper, we have leveraged real-world experimental datasets to evaluate the performance of position-aided mmWave V2I communications. We evaluated and compared the performance of three ML algorithms (i.e., KNN, SVM and DT) on mmWave V2I beam prediction using GPS coordinates as ML features. The results show the impact of codebook sizes on the accuracies of the ML algorithms under different scenarios. The results also reveal the limitations of beam prediction using geolocation side information only as average accuracy in some scenarios are less than 30%. Such low performances

underscore the difference between using synthetic and real-world experimental data. Therefore, mmWave V2I will benefit from a combination of sensors' data to enhance prediction accuracy results. Future research works will consider other ML and DL approaches as well as multi-modal beam prediction approach.

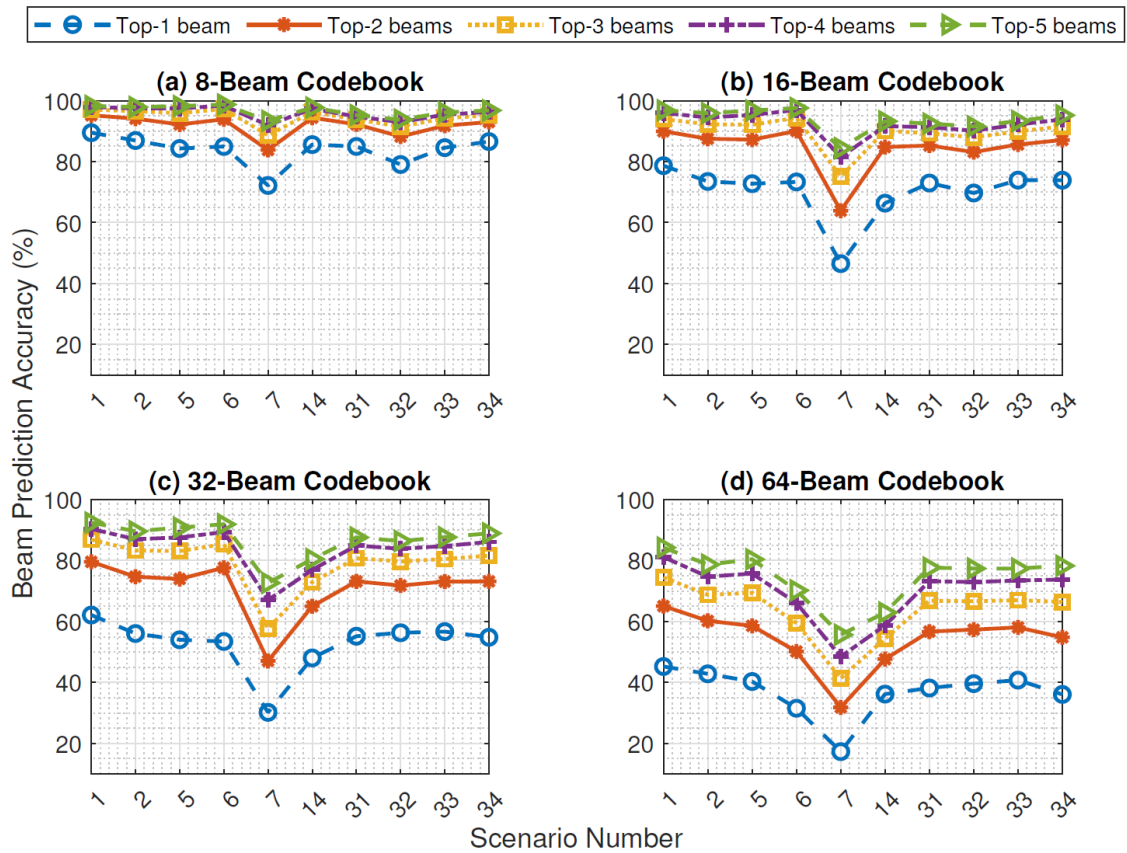


Figure 7. Top-k Beam Prediction Accuracy for $Q = 8, 16, 32$ and 64 using KNN

CRedit Author Contribution Statement

Sherif Adeshina Busari: Conceptualisation, Writing – original draft, Formal analysis, Investigation, Methodology, Funding acquisition; Ifiok Otung: Writing – review and editing, Supervision, Funding acquisition; Muhammad Ali: Writing – review and editing, Formal analysis, Methodology, Supervision; Raed Abd-Alhameed: Writing – review and editing, Supervision, Project administration, Funding acquisition.

Acknowledgement

This work is supported by the EPSRC [grant number EP/Z001544/1] through the UKRI-funded MSCA Postdoctoral Fellowship. It is also partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/Y035135/1, and HORIZON-MSCA-2022-SE-01-01-ID: 101131501.

References

- [1] Sherif Adeshina Busari, Shahid Mumtaz, Saba Al-Rubaye and Jonathan Rodriguez, "5G Millimeter-Wave Mobile Broadband: Performance and Challenges", *IEEE Communications Magazine*, Print ISSN: 0163-6804, Electronic ISSN: 1558-1896, Vol. 56, No. 6, pp. 137–143, 18 June 2018, Published by Institute of Electrical and Electronics Engineers (IEEE), DOI: 10.1109/MCOM.2018.1700878, Available: <https://ieeexplore.ieee.org/document/8387217>.
- [2] Davi da Silva Brilhante, Joanna Carolina Manjarres, Rodrigo Moreira, Lucas de Oliveira Veiga, José F. de Rezende *et al.*, "A Literature Survey on AI-Aided Beamforming and Beam Management for 5G and 6G Systems", *Sensors*, ISSN: 1424-8220, Vol. 23, No. 9, 28 April 2023, Published by Multidisciplinary Digital Publishing Institute (MDPI), DOI: 10.3390/s23094359, Available: <https://www.mdpi.com/1424-8220/23/9/4359>.

- [3] Qing Xue, Chengwang Ji, Shaodan Ma, Jiajia Guo, Yongjun Xu *et al.*, "A Survey of Beam Management for mmWave and THz Communications Towards 6G", *IEEE Communications Surveys and Tutorials, Electronic*, ISSN: 1553-877X, Vol. 26, No. 3, pp. 1520–1559, 5 February 2024, Published by IEEE, DOI: 10.1109/COMST.2024.3361991, Available: <https://ieeexplore.ieee.org/document/10422712>.
- [4] Sherif Adeshina Busari, Muhammad Awais Khan, Kazi Mohammed Saidul Huq, Shahid Mumtaz and Jonathan Rodriguez, "Millimetre-wave massive MIMO for cellular vehicle-to-infrastructure communication", *IET Intelligent Transport Systems*, Print ISSN: 1751-956X, Online ISSN: 1751-9578, Vol. 13, No. 6, pp. 983–990, 25 January 2019, Published by Institution of Engineering and Technology (IET), DOI: 10.1049/iet-its.2018.5492, Available: <https://digital-library.theiet.org/doi/full/10.1049/iet-its.2018.5492>.
- [5] Junyi Wang, Zhou Lan, Chin-Sean Sum, Chang-Woo Pyo, Jing Gao *et al.*, "Beamforming Codebook Design and Performance Evaluation for 60GHz Wideband WPANs", in *Proceedings of the IEEE 70th Vehicular Technology Conference Fall (IEEE VTC-Fall 2009)*, 20-23 September 2009, Anchorage, AK, USA, Print ISBN: 978-1-4244-2514-3, Print ISSN: 1090-3038, pp. 1-6, Published by IEEE, DOI: 10.1109/VETECF.2009.5379063, Available: <https://ieeexplore.ieee.org/document/5379063>.
- [6] João Morais, Arash Bchboodi, Hamed Pezeshki and Ahmed Alkhateeb, "Position-Aided Beam Prediction in the Real World: How Useful GPS Locations Actually are?", in *Proceedings of the IEEE Conference on Communications*, 28 May 2023 - 01 June 2023, Rome, Italy, Electronic ISBN: 978-1-5386-7462-8, Electronic ISSN: 1938-1883, pp. 1-6, Published by IEEE, DOI: 10.1109/ICC45041.2023.10278998, Available: <https://ieeexplore.ieee.org/document/10278998>.
- [7] Qiyu Duan, Taejoon Kim, Huang Huang, Kunpeng Liu and Guangjian Wang, "AoD and AoA tracking with directional sounding beam design for millimeter wave MIMO systems", in *Proceedings of the IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 30 August 2015 - 02 September 2015, Hong Kong, China, Electronic ISBN: 978-1-4673-6782-0, pp. 1-6, Published by IEEE, DOI: 10.1109/PIMRC.2015.7343676, Available: <https://ieeexplore.ieee.org/document/7343676>.
- [8] Marius Arvinte, Marcos Tavares and Dragan Samardzija, "Beam Management in 5G NR using Geolocation Side Information", in *Proceedings of the 53rd Annual Conference on Information Sciences and Systems (CISS)*, 20-22 March 2019, Baltimore, MD, USA, Electronic ISBN: 978-1-7281-1151-3, ISBN: 978-1-7281-1152-0, pp. 1-6, Published by IEEE, DOI: 10.1109/CISS.2019.8692820, Available: <https://ieeexplore.ieee.org/document/8692820>.
- [9] Chen Sun, Le Zhao, Tao Cui, Haojin Li, Yingshuang Bai *et al.*, "AI Model Selection and Monitoring for Beam Management in 5G-Advanced", *IEEE Open Journal of the Communications Society*, Electronic ISSN: 2644-125X, Vol. 5, pp. 38–50, 30 November 2023, Published by IEEE, DOI: 10.1109/ojcoms.2023.3337850, Available: <https://ieeexplore.ieee.org/document/10335766>.
- [10] Yuqiang Heng and Jeffrey G. Andrews, "Machine Learning-Assisted Beam Alignment for mmWave Systems", in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, 09-13 December 2019, Waikoloa, HI, USA, Electronic ISBN: 978-1-7281-0962-6, ISBN: 978-1-7281-0963-3, Electronic ISSN: 2576-6813, pp. 1-6, Published by IEEE, DOI: 10.1109/GLOBECOM38437.2019.9013296, Available: <https://ieeexplore.ieee.org/document/9013296>.
- [11] Sajad Rezaie, Elisabeth de Carvalho and Carles Navarro Manchón, "A Deep Learning Approach to Location- and Orientation-Aided 3D Beam Selection for mmWave Communications", *IEEE Transactions on Wireless Communications*, Print ISSN: 1536-1276, Electronic ISSN: 1558-2248, Vol. 21, No. 12, pp. 11110 - 11124, 18 July 2022, Published by IEEE, DOI: 10.1109/TWC.2022.3189788, Available: <https://ieeexplore.ieee.org/document/9832551>.
- [12] Yuyang Wang, Aldebaro Klautau, Mónica Ribero, Anthony C. K. Soong and Robert W. Heath "MmWave Vehicular Beam Selection with Situational Awareness Using Machine Learning", *IEEE Access*, Electronic ISSN: 2169-3536, Vol. 7, pp. 87479–87493, 10 June 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2922064, Available: <https://ieeexplore.ieee.org/document/8734054>.
- [13] Ahmed Alkhateeb, Gouranga Charan, Tawfik Osman, Andrew Hredzak, Joao Morais *et al.*, "DeepSense 6G: A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset", *IEEE Communications Magazine*, Print ISSN: 0163-6804 Electronic ISSN: 1558-1896, Vol. 61, No. 9, pp. 122–128, 5 June 2023, Published by IEEE, DOI: 10.1109/MCOM.006.2200730, Available: <https://ieeexplore.ieee.org/document/10144504>.
- [14] Karamot Kehinde Biliaminu, Sherif Adeshina Busari, Joaquim Bastos and Jonathan Rodriguez, "Impacts of Dataset and Codebook Sizes on ML-Driven Beam Prediction for mmWave V2I Communication", in *Proceedings of the IEEE 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 23-25 July 2024, Leeds, United Kingdom, Electronic ISBN: 979-8-3503-7786-6, ISBN: 979-8-3503-7787-3, Electronic ISSN: 2769-9994, pp. 1-6, Published by IEEE, DOI: 10.1109/WINCOM62286.2024.10655630, Available: <https://ieeexplore.ieee.org/document/10655630>.
- [15] Karamot Kehinde Biliaminu, Sherif Adeshina Busari, Jonathan Rodriguez and Felipe Gil-Castiñeira, "Beam Prediction for mmWave V2I Communication Using ML-Based Multiclass Classification Algorithms", *Electronics*, ISSN: 2079-9292, Vol. 13, No. 13, 6 July 2024, Published by MDPI, DOI: 10.3390/electronics13132656, Available: <https://www.mdpi.com/2079-9292/13/13/2656>.

- [16] Mohammed Shafi Kundiladi, Sheik Masthan Shahul Abdul Rahim and Mohammed Shahal Rishad, "Secure Autonomous Vehicle Localization Framework using GMCC and FSCH-KMC under GPS-Denied Locations", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol. 8, No. 3, pp. 64-74, 1 July 2024, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2024.03.005, Available: <http://aetic.theiaer.org/archive/v8/v8n3/p5.html>.
- [17] Norlina Mohd Sabri and Siti Fatimah Azzahra Hamrizan, "Prediction of MUET Results Based on K-Nearest Neighbour Algorithm", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol. 7, No. 5, pp. 50-59, 5 October 2023, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2023.05.005, Available: <http://aetic.theiaer.org/archive/v7/v7n5/p5.html>.
- [18] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua and Asdrubal Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends", *Neurocomputing*, ISSN 0925-2312, Vol. 408, pp. 189-215, 30 September 2020, Published by Elsevier, DOI: 10.1016/j.neucom.2019.10.118, Available: <https://www.sciencedirect.com/science/article/abs/pii/S0925231220307153>.
- [19] John C. Platt, Nello Cristianini and John Shawe-Taylor, "Large Margin DAGs for Multiclass Classification", in *Proceedings of the 12th International Conference on Neural Information Processing Systems (NIPS'99)*, 29 November 1999, Cambridge, MA, USA, pp. 547-553, Published by MIT Press, DOI: 10.5555/3009657.3009735, Available: <https://dl.acm.org/doi/10.5555/3009657.3009735>.
- [20] Johan Note and Maaruf Ali, "Intrusion Detection System Using Machine Learning and Deep Learning Algorithms", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol. 6, No. 3, pp. 19-36, 1 July 2022, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2022.03.003, Available: <http://aetic.theiaer.org/archive/v6/v6n3/p3.html>.
- [21] Bahzad Taha Jijo and Adnan Mohsin Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning", *Journal of Applied Science and Technology Trends*, ISSN: 2708-D787, vol. 2, no. 1, pp. 20-28, 24 March, 2021, Published by Interdisciplinary Publishing Academia, DOI: 10.38094/jastt20165, Available: <https://jastt.org/index.php/jasttpath/article/view/65>.



© 2025 by the author(s). Published by Annals of Emerging Technologies in Computing (AETiC), under the terms and conditions of the Creative Commons Attribution (CC BY) license which can be accessed at <http://creativecommons.org/licenses/by/4.0>.