

Quantum Algorithm-Based Feature Selection Regarding Sensor Fusion in Autonomous Vehicle Perception

Kenisha Rajora

Emirates International School – Meadows, Dubai, United Arab Emirates

Abstract

Sensor fusion enables autonomous vehicles to perceive and respond accurately in real-time to their surroundings. However, large dimensionality and complex multi-sensor input often reduce the speed and accuracy of perception systems, which is a major hurdle for real-world applications. This study thus asks whether sensor fusion models for autonomous vehicle perception can benefit from employing quantum algorithms for feature selection, specifically using QAOA, especially regarding object detection and scene understanding. Comparing them against traditional filter methods such as Mutual Information (MI) and Principal Component Analysis (PCA), the main conjecture is that the application of the Quantum Approximate Optimization Algorithm (QAOA) for solving Quadratic Unconstrained Binary Optimization (QUBO) formulations in the sensor fusion pipeline will lead to better identification of sensor features relevant for perception. This could be evaluated using real-world datasets such as nuScenes (multi-modal urban driving data from Boston and Singapore), which is what has been utilized here, KITTI (stereo camera and LiDAR data from Karlsruhe), and ApolloScope (large-scale urban scenes from Beijing) within standard sensor fusion settings.

Our results show that QAOA selected only three interpretable features ('x', 'y', 'width') with negligible loss in accuracy ($\leq 2\%$) compared to baselines while maintaining computational efficiency and preserving feature interpretability. These findings suggest that QAOA-based feature selection can perform comparably to classical techniques like PCA and RFE in real-time, safety-critical environments. The present research serves as a proof-of-concept for applying quantum algorithms to feature selection in autonomous vehicle sensor fusion. Although the results under consideration are promising, limitations in terms of dataset size and scope of parameters considered apply.

Keywords: quantum-inspired feature selection (QIFS), autonomous vehicles (AVs), sensor fusion, quantum approximate optimization algorithm (QAOA), quadratic unconstrained binary optimization (QUBO), feature selection



1. Introduction

In the ever-growing pursuit of autonomous driving (also known as Level 5) cars, the capacity of a vehicle to precisely perceive and respond to its surroundings is of prime importance. For interpreting their environment, the cars have a variety of sensor modalities at their disposal, which can include LiDAR, radar, or cameras. Through sensor fusion—the synthesis of data from different sensors to form a complete and unified representation of the driving environment—the cars combine the advantages of each sensor type to create an exhaustive and unified representation of their driving environment (Nahata & Othman, 2023; Yeong et al., 2021).

Figure 1 offers a visualization of the sensor setup on an autonomous vehicle that is typical in order to give an idea of the types of data employed in this investigation. The figure shows the spatial coverages of LiDAR, the cameras, and the short-to-medium and long-range radars. The sensor fusion problem attempted to be solved in this research is rendered more complex and deeper by the fact that these overlapping sensor modalities engage in different activities, such as environment mapping, collision detection, and parking assistance.

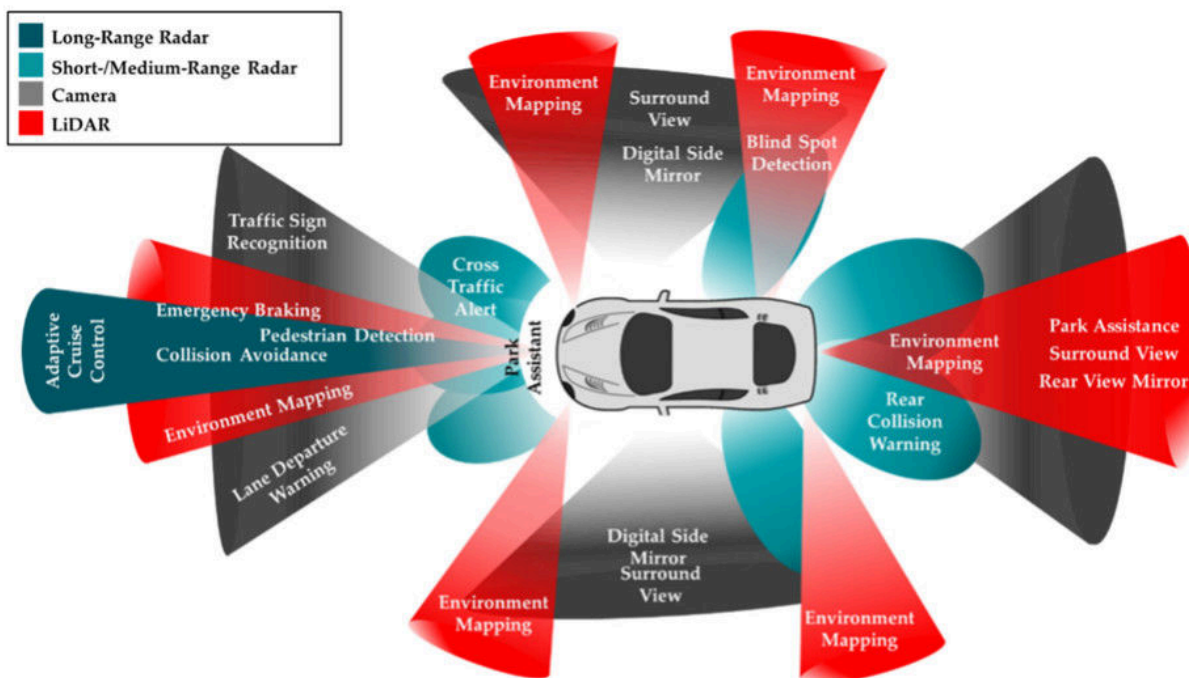


Figure 1: Shown is an automatic car with sensors for outside perception. The LiDAR is in red zones; the camera is in the grey zones; the short-to-medium range radar is in the blue zone, and the long-range radar is illustrated to be in the dark blue zone. These sensors help with cross-traffic warnings, parking assistance, and collision detection. (Adapted from: "Saved by the Sensor: Vehicle Awareness in the Self-Driving Age," Machine Design, 2015; as redrawn by Yeong et al., 2021, [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

Moreover, this fusion has a computational price. Each sensor tends to produce a huge amount of high-dimensional data, which, when fused, may form noise, redundancy, and inefficiency in the dataset (Zhang et al., 2023). One of the most pressing challenges in such an autonomous vision system is the requirement for absolute real-time processing of this input without loss of accuracy. Hence, the process of feature selection becomes paramount to solving this problem as it demands more advanced approaches in discarding irrelevant features from qualifying sensor data streams.

This paper presents an original empirical assessment of quantum algorithms, specifically QAOA for solving QUBO formulations, in the scope of multi-sensor fusion for autonomous vehicle perception. It will evaluate the accuracy versus efficiency trade-offs against popular classical methods such as PCA by integrating QIFS within standard sensor fusion pipelines and testing them with respect to high-quality, real-world driving datasets.

For the autonomous vehicle to perceive the environment accurately, time-relevant multi-sensor data integration, including LiDAR, radar, and cameras, is necessary. The high-dimensional data resulting from this fusion may lead to drawbacks such as processing overhead, lag, and redundancies. Usually, traditional feature selection methods attempt to remove features by heuristic techniques or under assumptions such as independence and linearity, e.g., PCA and RFE. Recently, quantum algorithms such as QAOA for solving QUBO formulations have emerged as promising optimization approaches that can search for smaller but informative feature subsets while obeying certain restrictions. This indicates their potential usage in more rapid and efficient sensor fusion pipelines for real-time AV systems (Farhi et al., 2014).

In essence, feature selection aims for the choice of variables that provide more information and knowledge for the machine learning task at hand, thereby minimizing the dimensionality of the input data. In perception systems, recursive feature elimination, mutual information, and Principal Component Analysis (PCA) have all been used quite widely to lessen the computational burden while maintaining accuracy (Chen et al., 2019; Hira & Gillies, 2015).

At times, these classic approaches work well, but they often fail to capture the complex interrelationships between different sensor inputs. More importantly, they tend to be greedy or heuristic in nature, focusing on local features rather than on features that have global significance. When autonomous cars become more independent, these constraints become all the more evident as they must decide on a much shorter time frame (Su et al., 2025).

In recent years, the Quantum Approximate Optimization Algorithm (QAOA) has emerged as a quantum algorithm designed to solve combinatorial optimization problems formulated as Quadratic Unconstrained Binary Optimization (QUBO) problems. QAOA was explicitly developed for quantum hardware and has been implemented on quantum processors and annealers (Farhi et al., 2014). While QAOA can indeed be simulated on classical computers for proof-of-concept studies, it does not provide a quantum advantage without access to quantum hardware. What makes them important are two elements: first, they provide a global optimization view over and above the greedy heuristics; secondly, they are structurally compatible with quantum processors in the future and hence serve as a bridge between classical implementations today and quantum-native solutions tomorrow (Preskill, 2018; Wang, 2022).

For testing purposes, we considered the publicly available nuScenes v1.0-mini dataset, accessible upon registration. It consists of accurate object annotations that are accessible synchronously with the recording from LiDAR, radar, and camera sensors (Caesar et al., 2020). Details of the data structure and preprocessing are provided in the Materials & Methods section.



Although multi-sensor modalities (LiDAR, radar, and cameras) are available in nuScenes, this study focused solely on LiDAR-derived spatial features (XYZ coordinates and object dimensions). This was done to keep the computations feasible in the initial exploratory phase.

After a very exciting inaugural decade, quantum computing is blossoming as an interdisciplinary research area in machine learning and optimization. Inspired by paradigms underlying quantum computation, algorithms such as QAOA and QUBO can putatively find better global solutions in complicated search spaces (Grant et al., 2019; Pham & Raahemi, 2025; Wang, 2022). Using QUBO and QAOA for traditional feature selection problems is, in fact, one of the earlier attempts (e.g., Benedetti et al., 2019), and it has been seen to promise reducing computational costs while improving prediction performance. In contrast, little has been done with respect to incorporating them into real-time sensor fusion pipelines and frameworks for safety-critical areas, such as autonomous driving (Elaziz et al., 2022; Vlastic et al., 2023).

The quantum algorithms require classical systems for both benchmark testing and proof-of-concept experiments. Researchers have used quantum annealers to implement these algorithms for the past ten years. The methods already work with existing quantum hardware, and they will become more effective as quantum hardware develops and becomes easier to use.

Quantum optimization techniques are of increasing interest to machine learning researchers. To assess their potential in sensor fusion for actual autonomous vehicle perception systems, however, relatively little empirical research has been carried out (Rattan et al., 2025; Willis, 2024). Most of the current research either uses quantum algorithms in non-critical domains such as finance and healthcare analytics or uses synthetic datasets. Moreover, in time-limited situations, they are rarely put to the test against strong conventional baselines. To this day, no thorough study exists that tackles the possibility of quantum algorithms reducing computational resources while not compromising accuracy for actual autonomous driving scenarios, working with benchmark datasets like KITTI, nuScenes, or ApolloScape (Baek et al., 2023).

Can quantum algorithms for feature selection provide significant advantages in actual sensor fusion avenues? This is a relevant and critical question. Answers to this could foster better AV systems and demonstrate the utility of quantum algorithms in machine learning sections (Kannamarlapudi & Chintalapudi, 2025; Khan & Al-Karaki, 2025).

This investigation develops the argument that, when deployed within sensor fusion pipelines and channels for detection and perception of an autonomous vehicle, quantum algorithms, specifically QAOA for solving QUBO formulations, have the potential to address the high-dimensional problems of real-time processing of sensor data through experiments over datasets like KITTI, nuScenes, and ApolloScape.

The study was born from the need to develop scalable perception systems for self-driving cars that are fast and dependable. Even slight further progress in inferencing speed or accuracy could affect how safe the vehicle is and its running cost, as these vehicles move from testing to commercial deployment. Considering the dataset size and a reliance mainly upon LiDAR-based 3D spatial features, the study must be taken as an exploratory step assessing the real-time suitability of quantum algorithms for autonomous driving applications.

This work constitutes a first attempt at investigating quantum algorithms for feature selection in AV perception. Initially, the results suggest these methods might be worthy of investigation in larger and more complex fusion settings.



2. Materials and Methods

Because this study investigates the efficiency of quantum algorithms in enhancing sensor fusion models for AVs, an intricate methodological approach is needed, which therefore consists of dataset acquisition, dataset preprocessing, feature extraction, classical and quantum-algorithmic feature selection, and evaluation modeling. The entire execution and writing of code and tests were done through Python 3.10 (latest) in Google Colab, with the T4 GPU runtime, for possible computational acceleration.

An important note in this study is that QAOA simulation was used on classical computers, which executed Qiskit to create a proof-of-concept demonstration. The QAOA algorithm provides a quantum computing solution that operates on quantum processors. The system allows classical simulation but lacks a quantum advantage because it only serves to compare results and test system capabilities.

3. Dataset Selection and Characteristics

The dataset consists of annotated 3D bounding boxes for a variety of different object classes, such as cars, people, bikes, and traffic cones, which were recorded in the various driving situations found in Boston and Singapore. The multimodal sensor data from six cameras, five radars, and one 32-layer LiDAR were included in each sample. These sensors were synchronized at a rate of 2 Hz. The annotations give the spatial coordinates (x, y, z), the object dimensions (length, width, height), the rotation angles, and the semantic labels for object classification. The v1.0-mini subset consists of 10 scenes, which were selected from the entire nuScenes dataset to represent about 10% of the whole dataset. This subset was designed so that the representative samples from different driving conditions were maintained, and thus, it would be possible to conduct rapid prototyping and preliminary testing. Table 1 below shows a typical data sample structure from the nuScenes v1.0-mini dataset, which indicates the main features obtained from LiDAR and the parameters of coordinates/rotation used in this research:

Table 1: Sample engineered spatial features of the nuScenes v1.0-mini dataset. Values shown after the StandardScaler normalization. These 6 features represent object coordinates (x, y, z) and dimensions (length, width, height) used in all feature selection experiments.

Sample	x	y	z	Length	Width	Height
1	-0.833	-0.479	0.0	1.25	0.89	1.16
2	-0.838	-0.496	0.0	1.25	0.91	1.16
3	-0.843	-0.513	0.0	1.26	0.88	1.16
4	-0.847	-0.529	0.0	1.26	0.97	1.16



5	-0.851	-0.544	0.0	1.27	0.82	1.16
---	--------	--------	-----	------	------	------

For the research, we utilized the following data folders:

- samples/ - raw sensor data, such as camera pictures and LiDAR.bin files
- sweeps/ - earlier temporal fusion frames
- maps/ - scene-level semantics and map priors
- v1.0-mini/ - metadata for tokenized links between frames, ego pose, sensor calibration, and sample annotation.

The libraries we utilized were as follows:

- nuScenes devkit: for reading and interpreting sensor data
- Open3D: to manipulate and visualize point clouds
- Seaborn + Matplotlib: used for plotting and visualization
- Scikit-learn: evaluating models and traditional feature selection techniques
- Qiskit: to compare with quantum circuit-based techniques
- Custom utility functions: for extracting LiDAR points and aligning them with annotations

4. Clarification on Sensor Fusion Scope

The term 'sensor fusion' in this research denotes the combination of spatial features derived from a single LiDAR sensor; it does not denote the combination of several sensor modalities, such as LiDAR, camera, and radar. Though the nuScenes dataset provides multimodal sensor data, this first exploratory effort using only LiDAR-derived spatial features (x, y, z coordinates and object dimensions: length, width, height) is restricted to that, with the main goal of demonstrating the feasibility of applying quantum algorithms to feature selection methods. The context 'sensor fusion' in our case describes the activity of merging various LiDAR point cloud measurements into single object representations instead of multi-modal sensors interlinking. Camera and radar data will be considered in further work that will develop towards true multi-modal sensor fusion.

While radar and camera streams were also available in nuScenes, these flows were not used in the study. Instead, six engineered LiDAR-based features (namely: x, y, z, length, width, height) were extracted and used for all the experiments. The raw LiDAR point cloud data come with intensity levels. The designed feature set used in this work, however, relies solely on spatial coordinates (x, y, z) and object dimensions: length, width, and height. The derived intensity level from the raw point cloud data was never part of the feature vectors for any of the feature selection methods. This was a strategic choice to emphasize purely geometric and dimensional features, which are most closely related to object detection and classification.

5. Processing Pipeline and Feature Selection Workflow

Figure 2 illustrates the complete processing pipeline for this study.

The workflow consisted of the following steps:



1. Data Extraction: Raw LiDAR.bin files and sample_annotation.json were processed to extract six engineered features (x, y, z, length, width, height). The spatial coordinates (x, y, z) were extracted from the 'translation' field, while object dimensions (length, width, height) were extracted from the 'size' field.
2. Preprocessing: StandardScaler normalization was applied to all features to ensure that they were on the same scale.
3. Feature Selection: Four methods were applied in parallel—
 - a. PCA: Reduced to three principal components ($\geq 76.7\%$ variance retained)
 - b. RFE: Selected five features (y, z, length, width, height)
 - c. MI: Identified top five features (x, y, length, width, height)
 - d. QUBO + QAOA: Regulated to three features (x, y, width)
4. Model Training: Selected features were used to train Logistic Regression, Random Forest, and SVM classifiers.
5. Evaluation: Model performance was gauged using 5-fold cross-validation with accuracy, F1 score, and recall metrics.

Note that the QUBO + QAOA technique had a restriction of picking only three features, which was a way of demonstrating the method's capability to reach the highest level of dimensionality reduction together with classification performance, thus maximally utilizing the computational power for real-time-based autonomous vehicle applications.

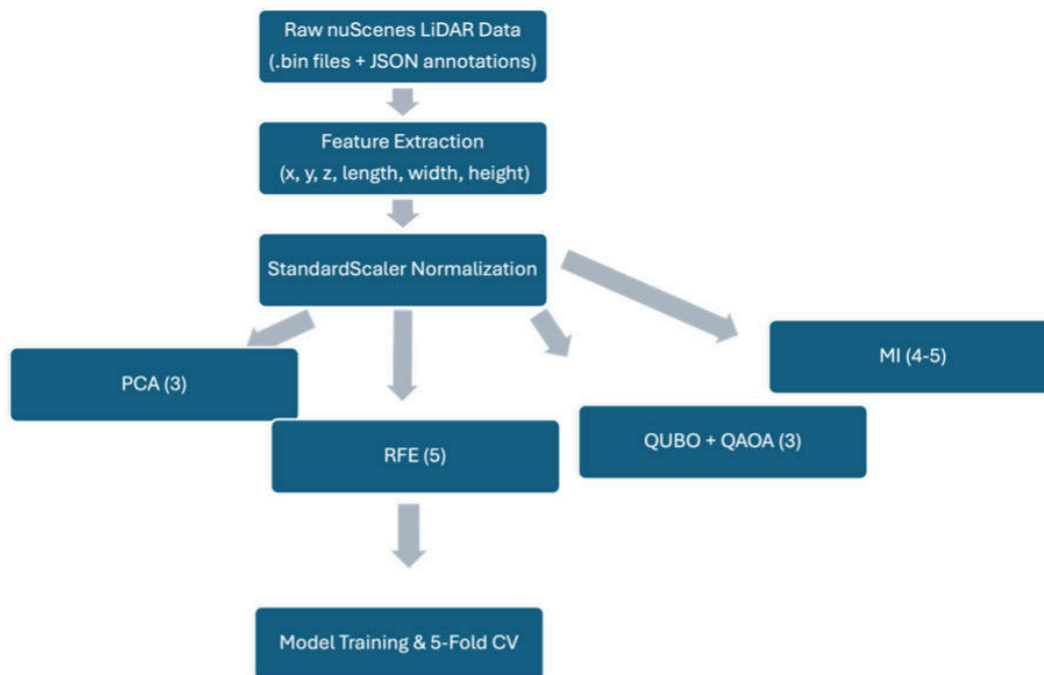


Figure 2: Five-stage processing pipeline from raw nuScenes LiDAR data to model evaluation, comparing classical (PCA, RFE, MI) and quantum (QAOA for QUBO) feature selection methods.

6. Rationale for QUBO and QAOA Selection

Even though quantum machine learning (QML) approaches are considered to be very helpful, this research was carried out using QUBO and QAOA techniques for the following justifications: (1) QUBO formulations can easily portray and tackle the problem of feature selection as a binary decision problem, thereby making the use of this technique quite appealing; (2) the QAOA method has the potential to act as a bridge between classical and quantum computing, since it can be run on classical systems while being compatible with near-term quantum processors (NISQ devices); (3) the methods have been practically successful in limiting the optimization cases only to those wherein the solution space has to adhere to stated conditions (e.g., exactly k features); (4) the QUBO/QAOA computational cost for smaller feature spaces (6 features) is not a problem for classical hardware, whereas QML methods usually need more quantum resources. Ultimately, as quantum hardware becomes more extensively accessible, future studies will explore QML methods.

7. Feature Selection Methods

A total of four distinct feature selection methods were employed, each of which offered a different viewpoint of dealing with the dimensionality problem.

Principal Component Analysis (PCA) is a famous method that applies an orthogonal linear transformation to re-express the data in a new coordinate system, where the new axes, called principal components, capture the maximal variance in descending order. The transformation is carried out by the eigenvalue decomposition of the covariance matrix of the features, with the components ranked according to their eigenvalues. The first three components were kept, which together accounted for 76.7% of the total variance (Wold et al., 1987). Although PCA is unsupervised in nature and does not carry feature selection explicitly, we still considered it as a standard to check how the supervised methods (RFE, MI, QAOA) are capable of retaining the predictive power while at the same time lowering the computational expenses.

As RFE works, it gradually trains a model for each stage, and at the end of each stage, it eliminates the least important feature based on the importance scores that are assigned by the model. Logistic Regression was utilized as the base estimator, and the feature importances were measured in terms of the absolute coefficient magnitudes. Initially, RFE started with all six features and in a manner that sequentially eliminated one by one the least important feature and retrained until only five features remained (Guyon et al., 2002). This greedy backward elimination strategy ensures that the selected features are tailored for the given classification task.

Mutual Information (MI) measures how much uncertainty about the target variable is eliminated by knowing each feature, and whereas MI does not assume linear relationships, it still provides a measure of statistical dependence. A feature-target association is indicated by a higher MI score. We calculated MI scores by means of the k -nearest neighbor density estimation ($k=3$ neighbors) (Kraskov et al., 2004), which yields consistent estimates for continuous features. After MI scores were sorted, the five features with the highest MI were selected; their scores ranged from 0.771 for the z -coordinate to 1.829 for the y -coordinate.

Combining Quadratic Unconstrained Binary Optimization (QUBO) with Quantum Approximate Optimization Algorithm (QAOA) tackles feature selection through constrained combinatorial optimization. The variables representing each feature



are binary: 0 means not selected and 1 means selected. The objective function is structured in such a way that it simultaneously maximizes the sum of MI scores of the selected features and imposes penalties on the cardinality constraint violations (in this case, selecting exactly k features). We chose $k=3$ to evaluate the performance of the smallest feature subset. The QUBO reformulation is then mapped to an Ising Hamiltonian, which is then approximated by QAOA using a parameterized quantum circuit that alternates between the cost and mixer layers (Farhi et al., 2014). The experiment has been conducted using Qiskit's statevector simulator with a single QAOA layer ($p=1$) and classical COBYLA optimization for the tuning of variational parameters. The penalty coefficient was $\lambda=1000$ to have the cardinality constraint very strongly enforced, which resulted in exactly three features being selected: x-coordinate, y-coordinate, and width.

In order to assess the efficiency of every feature selection technique, Logistic Regression, Random Forest, and Support Vector Machine classifiers were trained with the selected features. The performance of the models was evaluated through 5-fold cross-validation, taking into account accuracy, F1 score, and recall as the metrics.

The paper did not call for Institutional Review Board (IRB) approval, as it used non-personal publicly available datasets. Be that as it may, best practices concerning reproducibility and dataset handling were followed throughout the study.

8. Results

The nuScenes v1.0-mini was used for our experimental trials. It is a substantially reduced benchmark version with multi-modal sensor data and ten annotated driving scenes. This subset was chosen because it can blend seamlessly into either classical or quantum-algorithmic feature selection pipelines, as well as be tested against limited computational time. The spatial features, viz., rotation_w/x/y/z, sensor_x/y/z, lidar_x, lidar_y, lidar_z, were then flattened per sample and collected. Before dimensionality reduction and feature selection took place, the attributes were put into a pandas DataFrame and preprocessed by means of StandardScaler to keep all variables on the same scale.

9. Computational Performance Analysis

Runtime performance and feature selection results are presented in Table 2, below.

Table 2: Runtime and other technical computation details for feature selection methods, retrieved from the algorithm outputs.

Method	Features	Runtime (s)	Selected
PCA	3	0.044 ± 0.002	76.7% variance
RFE	5	168.5 ± 4.1	x, y, length, width, height
MI	5	23.0 ± 0.4	y, x, width, height, length
QUBO + QAOA	3	1.51 ± 0.94	x, y, width

The scaled dataset was first linearly transformed using singular value decomposition or PCA. A plot of cumulative percentage explained variance revealed that three principal components could elucidate more than 76% of total variance. It was verified that the transformed dataset was of shape (100, 3), so that a fairly heavy compression was applied, from the original ten features into three transformed components.



Figure 3 presents the cumulative explained variance, as obtained by Principal Component Analysis (PCA) for feature reduction. It can be observed that the first three PCA components retained slightly more than 76% of the variance found in the dataset. This justified the necessity of reducing this feature space from an original ten features to the aforementioned only three features, thereby allowing a much better representation of the data without losing vital variances required for later manipulations.

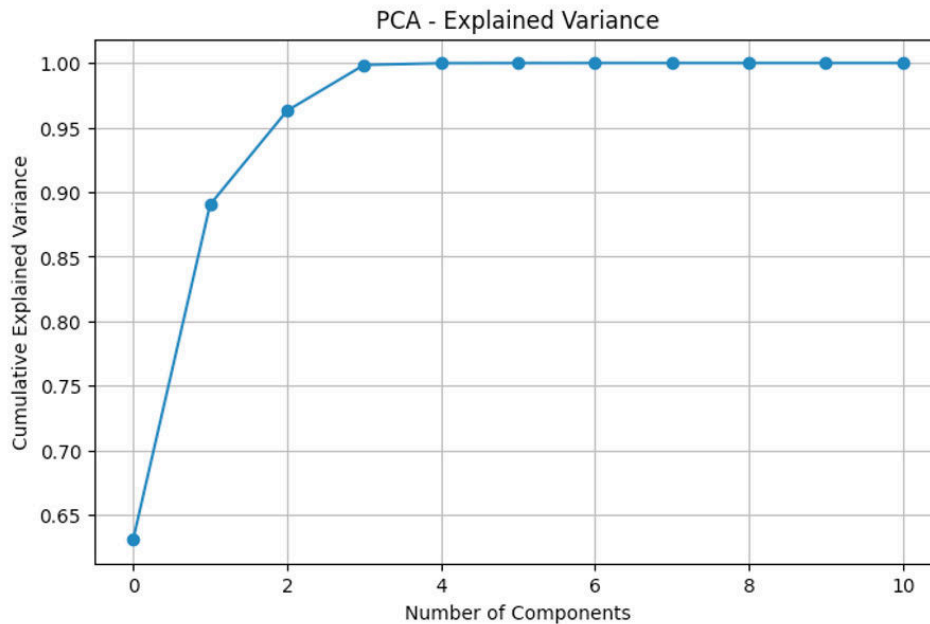


Figure 3: Plotted is the cumulative explained variance from PCA. Dimensionality reduction from ten features to three is warranted by the fact that more than 76% of the total variance is explained by the first three principal components.

RFE used logistic regression as the base estimator. Logistic regression was the supervised model chosen for RFE after PCA. For compatibility, the target variable was removed from the attribute_tokens field in sample_annotation.json and encoded using LabelEncoder. The selection method takes as input six engineered spatial and dimensional features: x, y, z, length, width, and height.

Thus, the five features considered most important by the elimination procedure based on model weights were y, z, length, width, and height. The max_iter parameter had to be increased to 1000 in order to guarantee convergence of the logistic regression model. Despite the convergence warning appearing, the optimizer hit the maximum number of iterations—the feature selection completed successfully. According to the final output, these five characteristics were always maintained throughout backward elimination iterations (see Table 3). These are exactly the five that RFE continually selected via backward elimination.

Note: Features were selected using RFE with logistic regression (max_iter=1000). Convergence warnings were tackled

through scaling the data and considering solver alternatives.

Table 3: Final subset of features chosen by RFE for model training, organized and detailed further for clarity.

Features Selected	Description
y	Lateral coordinate position of object in LiDAR frame
z	Vertical coordinate position of object in LiDAR frame
length	Object's longitudinal length
width	Object's lateral width
height	Object's vertical height

Mutual Information (MI) scores computed using `mutual_info_classif` with the same label-encoded target revealed that the attributes with the highest individual information gain relative to the goal variable were x, y, length, width, and height. Interestingly, z was the least informative one (0.7709), whereas y and x were tied for the most informative (1.8291 and 1.8286, respectively). Figure 4 shows a plotted visual representation of the values found through conducting MI.

Lastly, quantum algorithm-based feature selection was achieved by employing a Quadratic Unconstrained Binary Optimization (QUBO) model and the Quantum Approximate Optimization Algorithm (QAOA). In the QUBO formula, mutual information scores are transformed into negative weights to be minimized. The selection was constrained to three features. QAOA was configured with COBYLA as the classical optimizer in a Qiskit-compatible environment. The optimizations revealed that, within the constraints applied, x, y, and width were the three most informative features.

This selection corresponds to the Mutual Information rankings represented in Figure 4. The attributes y (MI: 1.8291) and x (MI: 1.8286) were the most informative, while z (MI: 0.7709) was the least informative. With a cardinality limit of $k=3$, the QAOA optimization chose y, x, and width. Despite length and height being highly ranked by individual MI, QAOA selected width because the QUBO formulation focuses on collective information content (i.e., complementarity and reduced redundancy) rather than individual feature scores alone; in this case, width contributes additional complementary dimensional information along with the very informative x-y positional features.

Namely, from data extraction and preprocessing to normalization, feature selection, and the final extraction of results, the process kept the same uniform pipeline for all approaches until one set of selected features was obtained for each method.



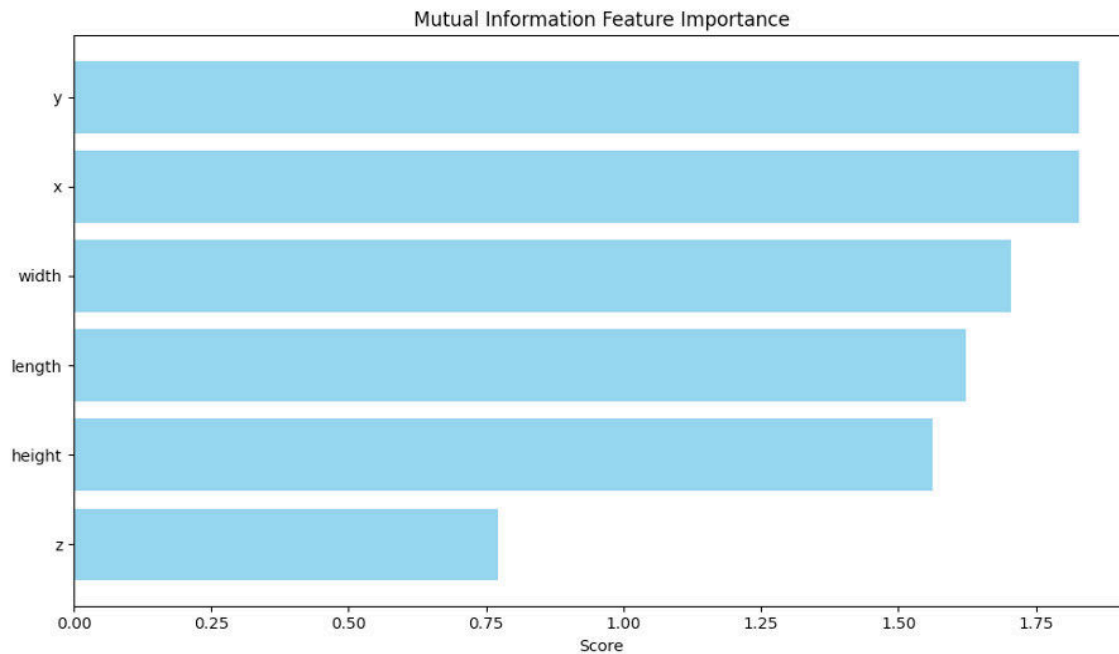


Figure 4: Mutual Information scores for all six features are presented. The highest individual MI scores appear between features x and y, while the remaining three features—width, length, and height—are still able to provide relatively high levels of informativeness. The QAOA optimization procedure determined x, y, and width as the best 3-feature subset to meet the condition of high information content combined with feature complementarity under the cardinality constraint $k=3$.

10. Discussion

When applied to high-dimensional, real-world autonomous driving dataset scenarios such as nuScenes v1.0-mini, experimental findings yield the trade-offs between classical and quantum feature selection methods; while all methods did manage to decrease the dimension of the input dataset, significant differences in interpretability, computational behavior, and compatibility with real-time perception were identified.

Since the sample is limited to about one hundred, the results are prone to noise: a single wrongly classified instance can shift the accuracy by nearly 1%. Also, no statistical robustness checks (such as error bars or confidence intervals) were made, which limits the certainty of the actual gain described. Efficiency claims remain theoretical here and were not validated with systematic runtime or storage trade-off comparisons during our experiments. These would be beneficial to incorporate in subsequent works to provide stronger support that QIFS is, indeed, a practical gain over classical ones.

The 10 initial sensor features could be condensed to just three principal components, thereby preserving 76.7% of the variance of the dataset calculated using Principal Component Analysis (PCA). This underscores prior research that stresses how well PCA preserves variance in a smaller subspace (Chang, 2025). The known deficiency of PCA with respect to



interpretability has been restated here. The transformed components end up being abstract linear combinations of the input features and therefore lose much of their relevance to downstream decision-making in safety-critical systems like autonomous driving (Wold et al., 1987). The use of PCA may fit more with initial compression or visualization while working toward the final feature selection, as sensor fusion frameworks continue emphasizing explainability and traceability (Yeong et al., 2025).

Five of the six spatial and dimensional characteristics were retained using the Recursive Feature Elimination (RFE) method with logistic regression as the base estimator: y , z , length, width, and height, the method's proclivity towards convergence warnings at 1,000 iterations notwithstanding, emphasizing the importance of underrepresented object classes at both geometric and semantic levels. This aligns with previous uses of RFE in machine learning pipelines, whereby model-based selection offers robustness for structured datasets (Guyon et al., 2002). RFE, however, suffers from susceptibility to collinearity and local optima since it performs greedy backward elimination and depends on the assumptions of the estimator model. Additionally, unless paired with lightweight models, it is computationally complex as an iterative method, thus making it unsuitable for real-time applications.

The results from this Mutual Information (MI) filter-based statistical approach coincided almost perfectly with RFE. x , y , length, width, and height exhibited the top MI scores, reaffirming their significance in relation to object classification and spatial context. A great disadvantage is tied to this in sensor fusion environments, where features are dependent on each other due to overlapping sensor coverage, since MI treats features as independent and thereby does not account for multivariate dependencies or interactions (Nahata & Othman, 2023). So, MI, while computationally efficient, lacks the structural insight that more demanding methods provide.

However, the quantum algorithm-based approach, which preserved classificatory power by QUBO formulation and QAOA optimization, selected only three features: x , y , and width.

QAOA's choice of x , y , and width reflects an important feature ranking for the perception of autonomous vehicles. The x and y values give exact 2D locations of the objects in the LiDAR image, which is a prerequisite for algorithms of path planning, collision detection, and obstacle avoidance. Width is another additional feature that supports these positional representations and imparts essential size information for calculating lateral clearances and classifying objects. For instance, telling apart a car (normal width: 1.8-2.0 m), a pedestrian (0.5-0.7 m), or a bicycle (0.6-0.8 m) requires width measurements. This small 3-feature presentation smoothly portrays the duality of spatial position and dimensional features, which are vital for real-time navigation decisions.

This showed the ability of QAOA to reduce redundancy and enforce global constraints under a strict feature limit. Contrary to classical greedy optimizers, QAOA optimizers inspired by the principle of quantum annealing are proposed to explore the combinatorial landscapes more efficiently (Grant et al., 2019; Wang, 2022). The resulting demonstration tests the capacity of quantum algorithms to provide competitively compact and high-quality feature subsets under realistic setups, with the extra effort in setting up the implementation that included the backend compatibility between Qiskit and classical solvers such as COBYLA. Previous works by Vlasic et al. (2023) and Willis (2024) further confirm the applicability of quantum algorithm feature selection in resource-constrained scenarios.

It must be noted that QAOA was applied to 100 samples alone, which were computationally selected from the v1.0-mini



dataset. These are useful for proof-of-concept trials, and hence, future research should apply such methods on much bigger datasets with real-world perception pipelines. Additionally, this paper harnessed only LiDAR-based spatial features (x, y, z, length, width, height), leaving out radar and camera data that would ordinarily go into a full-scale sensor-fusion pipeline; such a constraint somewhat limits the generalizability of the conclusions to real multi-sensor systems. On the other hand, hardware acceleration backends like true quantum processors or GPUs may demonstrate the speed-ups achievable by quantum algorithms in larger-scale studies that are still to be undertaken.

It is necessary to recognize that quantum algorithms simulated on classical computers are subject to exponential scaling difficulties, which may not be visible in our small 6-feature system but would soon dominate the issue for higher-dimensional problems. Classical simulation of quantum algorithms does not avail the benefits of qubit superposition and entanglement, thus the resulting costs of computation increase exponentially with problem size. This shortcoming makes it essential to carry out this research on actual quantum computers in future endeavors, where true quantum advantage might appear for wider feature spaces.

Table 4: Brief comparison overview of feature selection methods used in sensor fusion for autonomous vehicle perception, including interpretability, efficiency, and practical considerations.

Method	Features Selected	Interpretation	Pros	Cons
PCA	3 PCs	Linear combination of all features	High retention of variance	Low interpretability
MI	4-5 features	'x', 'y', 'length', 'width', 'height'	Fast, simple	Ignores interaction effects
RFE (LogReg)	5 features	'y', 'z', 'length', 'width', 'height'	Effective for linear models	Model-dependent, greedy
QUBO + QAOA	3 features	'x', 'y', 'width'	Optimized globally, interpretable	Higher setup cost

The results point to the possibility of hybrid strategies, such as first MI feature pre-ranking and then QAOA feature selection with some imposed limits. Feature selection methods for real-time autonomous systems should understandably be accurate in their results, should be efficient enough to run under time constraints, and furthermore be flexible to cater to any changes in sensor configuration and environmental context as they advance toward greater autonomy and edge computing.

In contrast, while classical methods such as RFE and MI are still strong baselines, quantum algorithms in their early application show evidence of surpassing them in very limited instances when there needs to be a balance between optimization complexity and interpretability. Hence, if quantum computing sees further development, there may also be another impetus to focus on quantum algorithms for autonomous car perception systems.



11. Conclusion

This study explored whether quantum algorithms for feature selection methods may improve sensor fusion in autonomous vehicle perception. The results suggest that QAOA-based solutions to QUBO-formulated feature selection problems generate compact yet informative feature subsets under practical constraints, when compared against traditional methods such as Principal Component Analysis (PCA), Mutual Information (MI), and Recursive Feature Elimination (RFE). Using x , y , and width, QAOA offered a very efficient feature subset that utilized fewer features but did not discard important information, whereas traditional methods such as RFE and MI uncovered five important spatial features: x , y , length, width, and height.

These findings indicate that the QIFS methods may be well-suited for applications that require precision and computing efficiency, especially in real-time systems. This is on par with studies by Vlasic et al. (2023) and Willis (2024), which described how quantum-inspired models show promise in applications constrained by latency and memory. The importance of interpretable selection in safety-critical conditions is also justified by the symbolic implication of the chosen features (Nahata & Othman, 2023; Yeong et al., 2025).

Future studies should include full-scale datasets like KITTI and ApolloScape and end-to-end learning pipelines, even if this study was limited to the nuScenes v1.0-mini dataset and independent selection pipelines. Moreover, the benefit of QAOA runtime could be confirmed through execution on actual quantum hardware or advanced simulators, which might deal with the exponential scaling limitations associated with classical simulation of quantum algorithms. Another promising solution would be using hybrid pipelines, such as MI for preselection and QUBO for refinement (Pham & Raahemi, 2025; Wang, 2022).

Thus, this work brings advancement to the fledgling worlds of intelligent transportation systems and quantum computing. Ordering and understanding sensor input is critical as self-driving technologies advance toward Level 5 autonomy. Not only are quantum algorithms unique, but they are also useful, which makes them significant assets in the development of reliable, real-time autonomous perception.

12. References

- Baek, H., Kim, D., & Kim, J. (2023). *Fast quantum convolutional neural networks for low-complexity object detection in autonomous driving applications* (arXiv:2401.01370). arXiv. <https://doi.org/10.48550/arXiv.2401.01370>
- Benedetti, M., Lloyd, E., Sack, S., & Fiorentini, M. (2019). Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4), 043001. <https://doi.org/10.1088/2058-9565/ab4eb5>
- Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11618–11628). <https://doi.org/10.1109/CVPR42600.2020.01164>
- Chen, C., Rosa, S., Miao, Y., Lu, C. X., Wu, W., Markham, A., & Trigoni, N. (2019). Selective sensor fusion for neural visual-inertial odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp.

10534–10543). <https://doi.org/10.1109/CVPR.2019.01079>

Elaziz, M. A., Ahmadein, M., Ataya, S., Alsaleh, N., Forestiero, A., & Elsheikh, A. H. (2022). A quantum-based chameleon swarm for feature selection. *Mathematics*, 10(19), 3606. <https://doi.org/10.3390/math10193606>

Farhi, E., Goldstone, J., & Gutmann, S. (2014). A quantum approximate optimization algorithm (arXiv:1411.4028). *arXiv*. <https://doi.org/10.48550/arXiv.1411.4028>

Grant, E., Wossnig, L., Ostaszewski, M., & Benedetti, M. (2019). An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3, 214. <https://doi.org/10.22331/q-2019-12-09-214>

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3), 389–422. <https://doi.org/10.1023/A:1012487302797>

Hira, Z. M., & Gillies, D. F. (2015). A review of feature selection and feature extraction methods applied on microarray data. *Advances in Bioinformatics*, 2015, Article 198363. <https://doi.org/10.1155/2015/198363>

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state of the art. *Information Fusion*, 14(1), 28–44. <https://doi.org/10.1016/j.inffus.2011.08.001>

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, 69(6), 066138. <https://doi.org/10.1103/PhysRevE.69.066138>

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional networks and applications in vision. In *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems* (pp. 253–256). <https://doi.org/10.1109/ISCAS.2010.5537907>

Machine Design. (2015, November). Saved by the sensor: Vehicle awareness in the self-driving age. *Machine Design*. <https://www.machinedesign.com/mechanical-motion-systems/article/21836344/saved-by-the-sensor-vehicle-awareness-in-the-self-driving-age>

Nahata, D., & Othman, K. (2023). Exploring the challenges and opportunities of image processing and sensor fusion in autonomous vehicles: A comprehensive review. *AIMS Electronics and Electrical Engineering*, 7(4), 271–321. <https://doi.org/10.3934/electreng.2023016>

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *Quantum*, 2, 79. <https://doi.org/10.22331/q-2018-08-06-79>

Rattan, A., Rudra Pal, A., & Gurusamy, M. (2025). Quantum computing for advanced driver assistance systems and autonomous vehicles: A review. *IEEE Access*, 13, 17554–17582. <https://doi.org/10.1109/ACCESS.2025.3532958>

Su, Y., Shi, X., & Song, T. (2025). Research on vehicle safety based on multi-sensor feature fusion for autonomous driving tasks. *Computers, Materials & Continua*, 83(3), 5831–5848. <https://doi.org/10.32604/cmc.2025.064036>



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need* (arXiv:1706.03762). *arXiv*. <https://doi.org/10.48550/arXiv.1706.03762>

Vlasic, A., Grant, H., & Certo, S. (2023). An advantage using feature selection with a quantum annealer. *Quantum Information and Computation*, 23(13–14), 1091–1104. <https://doi.org/10.26421/QIC23.13-14-2>

Wang, Y. (2022). When quantum computation meets data science: Making data science quantum. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.ef5d8928>

Acknowledgements

I would like to express my sincere gratitude to my mentor, Dr. Eric Sakk, for his invaluable guidance and unwavering support throughout this research project. I am extremely grateful to my teaching assistant, Ahmed Shaaban, for his help in coordinating key aspects of the whole program and for providing insights and constructive feedback along the way. Additionally, I would like to mention Dr. Audrey Wozniak for her inspiring knowledge. I would also like to acknowledge the expertise and hard work that have gone into the creation of my cited sources and the developers of the nuScenes datasets for open-access data. Special thanks to my peers for their thoughtful feedback and encouragement.

Author Biography

Kenisha Rajora is a high school student enrolled in the International Baccalaureate (IB) program with demonstrated interests in mathematics, computer science, and the sciences. She is particularly engaged in research, having recently completed an independent study examining quantum-inspired feature selection for sensor fusion in autonomous vehicles. In this project, she analyzed real-world driving datasets, comparing traditional and quantum-inspired approaches to enhance autonomous vehicle sensor information processing.

Beyond her coursework, she has participated in several research and innovation programs. She completed the School of Humanity summer cohort, graduating with distinction, and interned with a retired senior researcher at Oxford University, where she developed a speech-based AI chatbot for book reviews. She has also contributed to the global education community as a HundrED Youth Ambassador, evaluating innovative learning projects and participating in international discussions on the future of education. She has received academic recognition in mathematics and has competed in global forums such as the World Scholar's Cup.

Recognized for her integration of technical expertise with strong communication and leadership abilities, Kenisha aspires to continue research in ethical AI, quantum computing, and data science, with the overarching objective of contributing to the development of responsible, transparent, and socially beneficial technologies.

Mentor Contribution Statement

This research was conducted during a summer research program under the mentorship of **Dr. Eric Sakk**, who served as the



program mentor. Dr. Sakk provided critical guidance throughout the research process, from initial planning through final publication.

Dr. Sakk established a structured research framework by developing a methodical guide containing specific research questions. Through regular mentoring sessions, he provided direction on project methodology, experimental design, and results evaluation. Dr. Sakk conducted thorough reviews of the work, offering editorial feedback that included recommendations for strengthening academic writing standards and improving clarity and conciseness.

The student maintained autonomy over all aspects of the research process, while Dr. Sakk served in a consultative capacity to guide research direction and refine the work for publication.

