



# Improving Efficiency of Customer Requirements Classification on Autonomous Vehicle by Natural Language Processing

Hao Wang<sup>1</sup>, Asrul Adam<sup>1</sup> and Fengrong Han<sup>2</sup>

<sup>1</sup>College of Engineering, University Malaysia Pahang, Kuantan, Malaysia

<sup>2</sup>Faculty of Electrical & Electronics Engineering, University Malaysia Pahang, Pekan, Malaysia

Received 18 Feb. 2020, Revised 12 Apr. 2020, Accepted 26 Jun. 2020, Published 1 Nov. 2020

**Abstract:** Safety is critical for autonomous vehicle, therefore quality management system method is crucial for the risks that may impact human beings. Quality management system help identify customer requirements and finally meet them. Customer requirements also include other aspects that customers or stakeholders are most concerned. Although many researches on customer perception had been done, they do not include all aspect of the requirements toward autonomous vehicle. Furthermore, they are most in text format or will be transfer to text format that convenient to store and read. In front of the large amount text data, classifying them become time and costs consuming. The customer requirements on autonomous vehicle are summarized and allocated in different categories. The natural language processing method is applied in this paper. This method shows its efficiency on dealing with customer requirements. The results provide valuable reference for autonomous vehicle developer and top management.

**Keywords:** Customer Requirements, Autonomous Vehicle, Natural Language Processing, Quality Management

## 1. INTRODUCTION

Autonomous vehicle is a complex electronic-based system that controls all mechanical, electrical and other components. The functions and properties of an autonomous vehicle crucial to people's safety [1]. For example, it should brake on time in order to avoid collision with other vehicles or avoid hurt the pedestrians.

Most of vehicle makers implement quality management system like IATF16949, ISO26262 and even Automotive SPICE in their conventional and even autonomous vehicle process that ensure the quality of their products [2]. Quality management system also require identify customer requirements and deliver them to product development activities and then finally meet customer requirements. Therefore, customer requirements play an important role in the autonomous vehicle development process.

Customer not only require a safer product, but also care about the comfort for instance noise, vibration, temperature, indoor air quality [3]. Many researchers provide customer perception in different aspect. They are valuable for the autonomous vehicle development even though the customer perceptions are based on the forecasting of future options because there are very few

customers experienced. The knowledge about the public expectations could help estimate the important variables in order to forecast the effects on the society [4].

Customer requirements are text documents, engineers or managers have to transfer customer requirements into text format even they are originally not text. Therefore, researchers start to investigate classify them by Natural Language Processing (NLP) method that could improve the efficiency when compare with handling them manually [5]. Automatically gathering the data in operando and deliver them into the individual product as "Digital Twin" and manage them over the product life cycle is an important character of current concept: Industry4.0 [6].

## 2. OBJECTIVES

The objectives of this paper focus on getting comprehensive customer requirements on autonomous vehicle, finding an efficient way to classify customer requirements on autonomous vehicles according to defined categories, applying and comparing different algorithms.

### A. Getting Comprehensive Customer Requirements on Autonomous Vehicle

Previous paper more focus on single area – customer perception, did not separate them in to detailed classification. In order to get comprehensive requirements on autonomous vehicle as input to development process, authors collect customer requirements in different areas: Customer perception, function, environment, energy consumption, costs, privacy, security, safety, society, legal and ethical [7]-[11]. The result will show in following sections.

### B. Finding an Efficient Way to Classify Customer Requirements

Another objective is that providing the NLP method to classify customer requirements on autonomous vehicle. This method efficiently classifies the text format requirements in several seconds that could help companies reduce the costs of human power. This paper also aimed to shape a customer requirements identification and classification process that help regulate the delivery of customer requirements from the raw information to the quality functional deployment. It provides a possibility to build an automatic electronic system that could replace human power and improve the efficiency of autonomous vehicle development [5].

### C. Applying and Comparing Different Algorithms

By applying different algorithms, the accuracy of machine learning algorithms in NLP are compared in order to find the best solution [5].

## 3. METHOD

A convention vehicle development project includes project preparation, customer value, concept definition, product design, process design, product and process validation, supply chain build-up and start-up, and finally ramp-up. See Fig.1. From this figure we can see that Customer Value locate in the early stage of the product development process. Quality management system requires obtain customer requirements and finally satisfy customer, that means customer value proposition plays an important role in the value stream. It not only decides the product specification of the product, but also decides whether customer or market will accept it. These are vital factors for a company's business. As a complicated and life vital product, autonomous vehicle is obviously a new thing for most of the vehicle company. Collecting and analyzing customer requirements become more important.

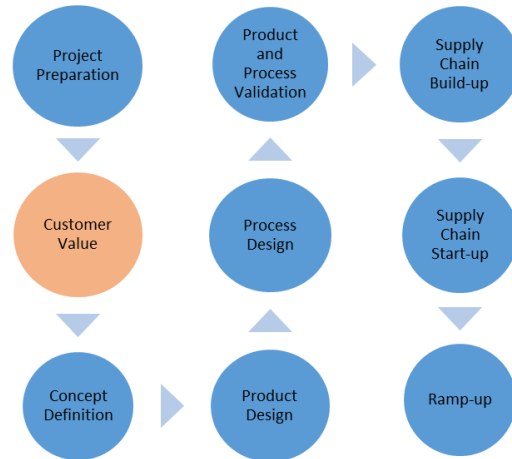


Figure 1. Convention Vehicle Development Process

In order to comprehensively collect and classify customer requirements, author provide the NLP method that use Python and NLTK to automate the process. The flow chart of this process shown in Fig.2. The steps of this flow chart are introduced in the following paragraphs of this Section.

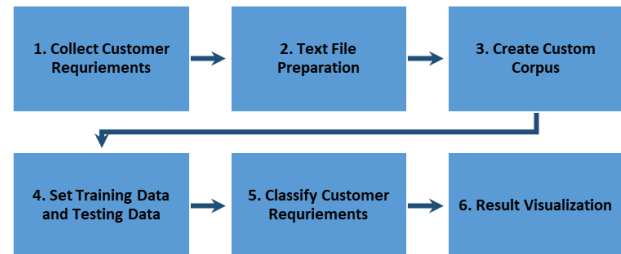


Figure 2. Customer Requirements Classification Flow Chart

### A. Collect Customer Requirements

Former researcher did very excellent job on collecting customer requirements. They deliver thousands of questionnaires to get the origin perception of autonomous vehicle from customers in different countries and districts.

A research on consumer demand for fully automated driving has been done in Japan. It evaluates consumers demand on purchase intention and willingness to pay for fully autonomous vehicle. The result shows that the major merits consumers care about are reduce the driving burden, get in and out of the car at desired locations, park automatically, and not need driver license. And the top demerits that the consumer care about are anxieties caused by autonomous vehicle, information leakage, and the car's availability [7]. Another similar research shows that the similar result on European citizens. It also considers the age of the citizen and cities' policies that could impact on the feeling of them about autonomous vehicles. Interestingly, many respondents show their responses about the usage of time when using an autonomous vehicle. They would be working, reading or relaxing [8].

One research selects fellow researcher as their respondents to investigate the customer perception of autonomous vehicles in order to draw implications for the policy makers and related stakeholders to facilitate market penetration. Engineers and managers in vehicle company who engaged in vehicle development process could benefit from the result. It identifies 30 promoters and 27 barriers that provide valuable insights for the product or service configuration an autonomous vehicle should have [9]. China’s public receptivity toward autonomous vehicle also evaluated by researchers. Drivers and non-drivers’ attitude were compared, it shows that drivers were concern less about autonomous vehicles. Finally, 18 benefits are concluded according to its autonomous vehicle acceptability scale [10]. Similar research also performed in Australia [11] and worldwide [12][13].

Although many researches on customer perception had been done, they do not include all aspect of the requirements toward autonomous vehicle. Therefore, author collect the research result from other aspects. They are from the aspect of stakeholders [14], legal and regulation [15], ethical [16], energy consuming [17] and costs [18], privacy and security [19][20][21], and car sharing [22].

Obviously, there are both positive and negative perception toward autonomous vehicle. From the vehicle development point of view, negative perceptions are also valuable for the product development. It plays the same role with the positive opinions. Therefore, in this paper, author classify the requirements according to its categories. Putting all these surveys result together could help make the requirements toward autonomous vehicle more comprehensive. Unavoidable there are many repeat items in the research results. But these repeated results become very useful data that help classify the customer requirements. Finally, picking up 144 customer requirement sentences from the papers and form a text format data base. The example of the customer requirements shown in Tab.I.

TABLE I. EXAMPLE OF CUSTOMER REQUIREMENTS

No.	Customer Requirements
1	Prone to hackers.
2	children can ride without supervision.
3	travel time savings.
4	Peace of mind.
5	A self-driving car would need less fuel.

**B. Text File Preparation**

After customer requirements collection, it is time to build the text file in order to create the customer corpus. In consider of the customer requirements has different categories, author split all raw requirements into six categories. And these categories are what we will classify in the following steps. The six categories are: 1. Environment, energy and costs; 2. Function; 3. Perception; 4. Privacy and security; 5. Safety; 6. Social,

legal and ethical. Six empty folders created with the simplified name according to the six categories. See Fig.3. The relations between the resources and the six categories please see Tab.II.

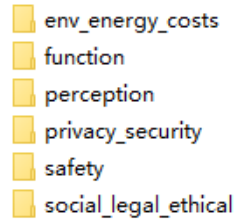


Figure 3. Create six folders with simplified category names

This step is quite time consuming because all sentences need to be identified and put into the six folders. It is valuable to spend a large amount of time on this job because it is the foundation of the customer corpus of natural language processing. Then we put each sentence into a txt format text file. See Fig.4. In consider of the customer requirements could be a paragraph, dealing with the data in this way convenient for future extending of customer requirements from internet forum or other resources.

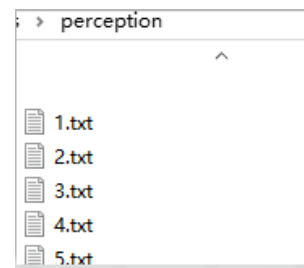


Figure 4. Put each sentence into a txt format text file(An example)

**C. Create Custom Requirements Corpus**

```

[('consumption': True, 'fuel': True, 'decrease': True),
 'env_energy_costs'], (('costs': True, 'cheaper': True,
 'transport': True, 'compared': True, 'modes': True,
 'transportation': True, 'public': True, 'taxi': True),
 'env_energy_costs'), (('self-driving': True, 'need': True,
 'fuel': True, 'would': True, 'less': True),
 'env_energy_costs'), (('self-driving': True,
 'vehicle': True, 'emissions': True, 'would': True,
 'lower': True), 'env_energy_costs'), (('self-driving': True,
 'rates': True, 'would': True, 'insurance': True,
 'lower': True), 'env_energy_costs'), (('increase': True,
 'costs': True, 'initial': True, 'maintenance': True),
 'env_energy_costs'), (('unpredictable': True,
 'risks': True, 'avoiding': True, 'financial': True),
 'env_energy_costs'), (('financial': True, 'interest': True),
 'env_energy_costs'),
 ...
 (('technologies': True, 'safety': True, 'general': True,
 'trust': True, 'worried': True), 'safety')]
    
```

Figure 5. Example of the custom corpus



Based on the text file created in last step, it is possible to create custom corpus. There are many exist corpus in NLTK but here in this research, they could not satisfy the needs of the research. Authors have to train our own model. Here we create a text classifier and custom corpus to train on. In fact, the easiest way to categorize the custom corpus is have one folder for each category. In last step, we create six folders, Read the six folder names as the name of categories. Then read the sentences as a preparation for tokenizer. Before tokenizing them, we

need to convert all upper case to lower case. The tokenizing process is a common pattern in NLP, it split a sentence into words. It provides a possibility to filter stop words. Comparing with the tokenizing process, it is also possible to stem and lemmatize the words that could improve the accuracy of classification. Furthermore, we also need get rid of relatively uninformative words that possible to impact on the accuracy. After above procedures, we finally create a custom corpus. See Fig.5.

TABLE II. CUSTOMER REQUIREMENTS MATRIX

Reference No.	Category					
	<i>Environment, Energy, Costs</i>	<i>Function</i>	<i>Perception</i>	<i>Privacy, Security</i>	<i>Safety</i>	<i>Social, Legal, Ethical</i>
[7]		✓	✓	✓	✓	✓
[8]	✓	✓	✓	✓	✓	✓
[9]	✓	✓	✓	✓	✓	✓
[10]	✓	✓	✓	✓	✓	✓
[11]	✓	✓	✓	✓	✓	✓
[12]		✓		✓	✓	✓
[13]		✓	✓		✓	
[14]						✓
[15]						✓
[16]						✓
[17]	✓					
[18]	✓					
[19]				✓		
[20]		✓		✓		✓
[21]		✓		✓		
[22]	✓	✓	✓	✓	✓	✓

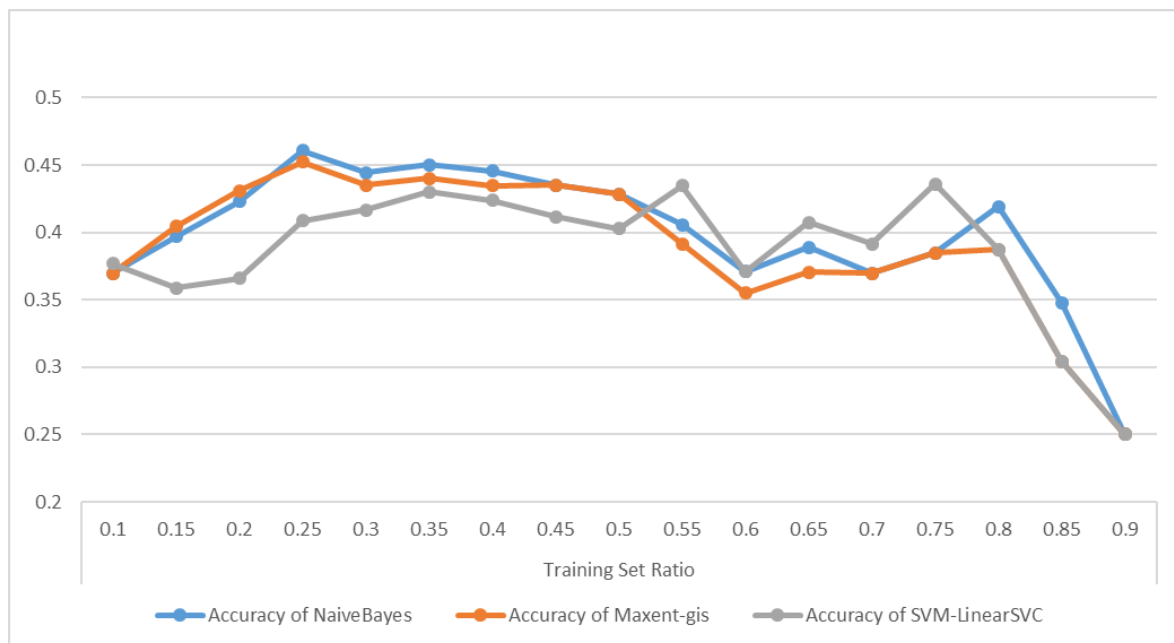


Figure 6. Accuracy of different classification algorithms under different training set ratio



TABLE III. ACCURACY OF DIFFERENT CLASSIFICATION ALGORITHMS UNDER DIFFERENT TRAINING SET RATIO

Algorithm	Training set ratio																
	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
NaiveBayes	0.37	0.40	0.42	0.46	0.44	0.45	0.45	0.44	0.43	0.41	0.37	0.39	0.37	0.38	0.42	0.35	0.25
Maxent-gis	0.37	0.40	0.43	0.45	0.44	0.44	0.43	0.44	0.43	0.39	0.35	0.37	0.37	0.38	0.39	0.30	0.25
SVM-LinearSVC	0.38	0.36	0.37	0.41	0.42	0.43	0.42	0.41	0.40	0.43	0.37	0.41	0.39	0.44	0.39	0.30	0.25

D. Set Training Data and Testing Data

After we create custom corpus, we have enough data to set training data and testing data. In order to reduce the impact on the accuracy we also need to mix the corpus. Then we split them into two parts, one for training and the rest for testing. We could set the ratio at 0.8, which means 80% for training and 20% for testing. The ratio could adjust according to the trial run results.

E. Classify Customer Requirements

Classification is the way to categorize documents or texts. Computer or classifier could decide between the labels or categories. We should try different classifier and algorithms in order to improve the accuracy of classification. In this paper, author use NaiveBayes, Maxent-gis and SVM-LinearSVC.

F. Result Visualization

In order to show and analyze the results, we should make the results into a more intuitive approach. The number is a simplest way, here in this paper, we more display and compare the accuracy in percentage points. In addition, we also display the matrix that show the deviation between the classification result and actual results after the testing.

4. RESULT

Following the steps shown in last chapter, we obtain the classification results. They are shown in Fig.6 and Tab.III. Author use accuracy to describe the quality of the classification. We can see that accuracy reach its top value 0.46 under training set ratio 0.25. Not only NaiveBayes classifier reach its highest number, the Maxent-gis also reach the highest number with accuracy 0.45. SVM-LinearSVC gain much lower accuracy, only 0.41. Interestingly, the lowest accuracy 0.25 appeared under the training set ratio 0.9. And the three algorithms gain the same value at this point. We also notice that the accuracy quite high under the training set ratio from 0.65 to 0.8. SVM-LinearSVC reach 0.44 at training set ratio 0.75, NaiveBayes reach 0.42 at training set ratio 0.8 and Maxent-gis reach 0.39 at training set ratio 0.8.

5. DISCUSSION

When checking the accuracy results in detail, we find that they are not reasonable. For example, it should gain higher accuracy if we use more training sets, the accuracy should lower if we use less training sets. Obviously, The result reversed from this point of view. Let's see what happened during the classification.

Accuracy: 0.460870  
Confusion Matrix:

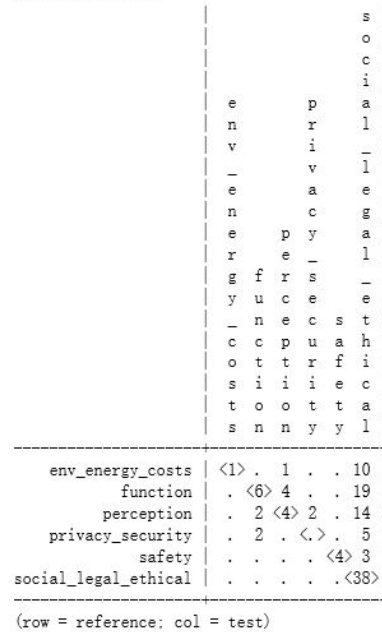


Figure 7. Confusion Matrix for Naïve-Bayes under training set ratio 0.25

Accuracy: 0.460870  
Confusion Matrix:

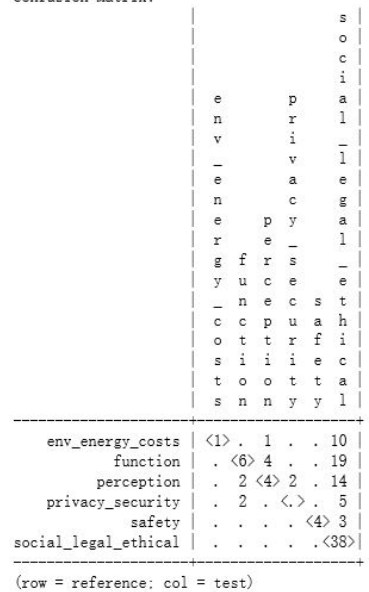


Figure 8. Confusion Matrix for Maxent-gis under training set ratio 0.25





Figure 9. Confusion Matrix for SVM-LinearSVC under training set ratio 0.25



Figure 10. Confusion Matrix for SVM-LinearSVC under training set ratio 0.75

Here we use confusion matrix. It not only provides the results inside each category, but also shows misclassified or not. The confusion matrix for NaiveBayes under training set ratio 0.25 shows that there is no correct classification for category of ‘privacy\_security’. And the correct result all accumulated in category of ‘Social\_Legal\_Ethical’. See Fig.7. It proves that many categories lack of training set and put most of the results

into the last category. The confusion matrix for Maxentgis under the same training set ratio shows the similar result. See Fig.8.

When we check SVM-LinearSVC, the result in confusion matrix much better. Each category with similar result. See Fig.9. The similar result shown under training set ratio 0.75. See Fig.10. From these data we can see that support vector machine algorithm show its advantage for classification in this paper.

The lowest accuracy 0.25 caused by there are no testing set for some categories when we set training set ratio at 0.9.

### 6. CONCLUSION

This paper provides a method and then creates corpus for customer requirements on autonomous vehicle. NLP method is used to classify the customer requirements. Three kinds of algorithms are compared in this research. SVM shows advantage in this research. This study illustrates that NLP helps classify the customer requirements on autonomous vehicle within several seconds. The inaccuracy somewhat caused by small volume of training set. The customer corpus could expand in the future in order to deal with bigdata from the website or server. Even the work in this paper is not perfect, it could still be a valuable reference for the researcher, manager and engineer who engaged in autonomous vehicle development.

### ACKNOWLEDGMENT

We are very grateful to previous researchers for their detailed investigation on consumer requirements on autonomous vehicle. Also thanks the researchers who help adjust the code on NLTK.

### REFERENCES

- [1] F. Wotawa, B. Peischl, F. Klück, and M. Nica, “Quality assurance methodologies for automated driving,” *Elektrotechnik und Informationstechnik*, vol. 135, no. 4–5, pp. 322–327, 2018.
- [2] S. Gifei and A. Salceanu, “Integrated Management System for quality, safety and security in developing autonomous vehicles,” *2017 10th Int. Symp. Adv. Top. Electr. Eng. ATEE 2017*, pp. 673–676, 2017.
- [3] I. Bae, J. Moon, and J. Seo, “Toward a Comfortable Driving Experience for a Self-Driving Shuttle Bus,” *Electronics*, vol. 8, no. 9, p. 943, 2019.
- [4] T. A. S. Nielsen and S. Haustein, “On sceptics and enthusiasts: What are the expectations towards self-driving cars?,” *Transp. Policy*, vol. 66, no. February, pp. 49–55, 2018.
- [5] A. Lyutov, Y. Uygun, and M. T. Hütt, “Managing workflow of customer requirements using machine learning,” *Comput. Ind.*, vol. 109, pp. 215–225, 2019.

- [6] C. Wunderlich, C. Tschöpe, and F. Duckhorn, "Advanced methods in NDE using machine learning approaches," AIP Conf. Proc., vol. 1949, no. April, 2018.
- [7] K. J. Shin, N. Tada, and S. Managi, "Consumer demand for fully automated driving technology," Econ. Anal. Policy, vol. 61, no. 2019, pp. 16–28, 2019.
- [8] O. Pribyl and M. Lom, "Impact of Autonomous Vehicles in Cities: User Perception," 2019 Smart City Symp. Prague, pp. 1–6, 2019.
- [9] K. Merfeld, M.-P. Wilhelms, and S. Henkel, "Being driven autonomously – A qualitative study to elicit consumers' overarching motivational structures," Transp. Res. Part C Emerg. Technol., vol. 107, no. January 2018, pp. 229–247, 2019.
- [10] W. Qu, J. Xu, Y. Ge, X. Sun, and K. Zhang, "Development and validation of a questionnaire to assess public receptivity toward autonomous vehicles and its relation with the traffic safety climate in China," Accid. Anal. Prev., vol. 128, no. April, pp. 78–86, 2019.
- [11] M. L. Cunningham, M. A. Regan, T. Horberry, K. Weeratunga, and V. Dixit, "Public opinion about automated vehicles in Australia: Results from a large-scale national survey," Transp. Res. Part A Policy Pract., vol. 129, no. August, pp. 1–18, 2019.
- [12] M. Kyriakidis, R. Happee, and J. C. F. De Winter, "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents," Transp. Res. Part F Traffic Psychol. Behav., vol. 32, pp. 127–140, 2015.
- [13] C. Tennant, S. Stares, and S. Howard, "Public discomfort at the prospect of autonomous vehicles: Building on previous surveys to measure attitudes in 11 countries," Transp. Res. Part F Traffic Psychol. Behav., vol. 64, pp. 98–118, 2019.
- [14] S. Pettigrew and S. L. Cronin, "Stakeholder views on the social issues relating to the introduction of autonomous vehicles," Transp. Policy, vol. 81, no. June, pp. 64–67, 2019.
- [15] C. Bartolini, T. Tettamanti, and I. Varga, "Critical features of autonomous road transport from the perspective of technological regulation and law," Transp. Res. Procedia, vol. 27, pp. 791–798, 2017.
- [16] N. Adnan, S. Md Nordin, M. A. bin Bahrudin, and M. Ali, "How trust can drive forward the user acceptance to the technology? In-vehicle technology for autonomous vehicle," Transp. Res. Part A Policy Pract., vol. 118, no. October, pp. 819–836, 2018.
- [17] Z. Ajanović, M. Stolz, and M. Horn, "A novel model-based heuristic for energy-optimal motion planning for automated driving," IFAC-PapersOnLine, vol. 51, no. 9, pp. 255–260, 2018.
- [18] V. Gružasuskas, S. Baskutis, and V. Navickas, "Minimizing the trade-off between sustainability and cost effective performance by using autonomous vehicles," J. Clean. Prod., vol. 184, pp. 709–717, 2018.
- [19] V. H. Le, J. den Hartog, and N. Zannone, "Security and privacy for innovative automotive applications: A survey," Comput. Commun., vol. 132, no. October, pp. 17–41, 2018.
- [20] R. de Bruin, "Autonomous intelligent cars on the european intersection of liability and privacy: Regulatory challenges and the road ahead," Eur. J. Risk Regul., vol. 7, no. 3, pp. 485–501, 2016.
- [21] S. Sharma and B. Kaushik, "A survey on internet of vehicles : Applications , security issues & solutions," Veh. Commun., vol. 1, p. 100182, 2019.
- [22] K. Merfeld, M. P. Wilhelms, S. Henkel, and K. Kreutzer, "Carsharing with shared autonomous vehicles: Uncovering drivers, barriers and future developments – A four-stage Delphi study," Technol. Forecast. Soc. Change, vol. 144, no. November 2017, pp. 66–81, 2019.



**Hao Wang** is currently a PHD candidate in College of Engineering, University Malaysia Pahang. His research interests include Autonomous Vehicle, Quality Management, Natural Language Processing, and Machine Learning. He has 10 years quality management experience in automotive industry and 5 years R&D and research experience.



**Dr. Asrul Adam** received the B.Eng. and M.Eng. degrees in Electrical-Mechatronics and Electrical Engineering from Universiti Teknologi Malaysia (UTM), Malaysia, in 2009 and 2012, respectively. He completed his Ph.D. in Signal and System from Faculty of Engineering, University of Malaya in 2017. He currently works as senior lecturer at College of Engineering, Universiti Malaysia Pahang, Malaysia. His research interests include biomedical signals processing, artificial intelligent, machine learning, and optimization.



**Fengrong Han** is currently working toward the Ph.D. degree in faculty of electrical & electronics engineering, University Malaysia Pahang, Malaysia. Her research interests include Localization in wireless sensor networks, Routing protocols, Multi-objective optimization, Signal processing, Intelligent system, Deep learning for Image and Signal classification. She has 3 years of software developing experience and 2 years of research experience.