

Comparison of Chi-Square Test and Representative Decision Tree in Features that Influence Vehicle Style

Hung-Hsiang Wang and Chih-Ping Chen

Abstract—This investigation tried to compare the features that influence vehicle style with Chi-square test and the representative decision tree method. This study has three outcomes. First of all, the investigation using Chi-square test could find evidence of the correlation between car style and some design features. Secondly, for the goal to improve accuracy, this investigation created a method of the representative decision tree. It built 50 decision trees to calculate accuracy to compare and to choose the best one tree. The third, although vehicle style was related to some design features, there were still differences in the chosen design features between the representative decision tree method and Chi-square test. The ranking of importance for the design features correlate with the vehicle style was not the same. Finally, we attempted to use design knowledge in this study to create a series of 3d modeling concepts with different vehicle design styles.

Index Terms—Representative decision tree, chi-square test, design features, vehicle style.

I. INTRODUCTION

Product aesthetics influenced product style. Take car design as an example which has a century history of development. The design history has recorded various styles in many periods. Trends and styles changed over time. Fortunately, with technological advancements, a collection of automotive design information would not be a problem. Whether we opt to collect data and photos from the yearbook or the internet, we can simply find some design knowledge from the past car database. Thus, we could discover the design features that influence design style.

Moreover, the method could be used in the stage of design and development to understand the design features of vehicle style. Let the designer or engineer create several new vehicles with various styles. Rather than thinking wildly and the designer's failure to grasp the design direction, this study provides a rational way.

This was also a multidisciplinary design study that combined industrial design, information science, and statistics to reasonably discuss vehicle styles. We could find design knowledge more effectively. It also used data mining to find some scientific evidence of the correlation between design features and historical styles.

Manuscript received January 15, 2020; revised November 25, 2020.

Hung-Hsiang Wang is with the Department of Industrial Design at National Taipei University of Technology, Taipei, Taiwan (e-mail: wanghh@mail.ntut.edu.tw).

Chih-Ping Chen is with the College of Design at National Taipei University of Technology, Taipei, Taiwan (Corresponding author; e-mail: roychen092@hotmail.com).

II. LITERATURE REVIEW

A. Artificial Intelligence and Data Mining

Artificial intelligence includes different topics such as data mining, machine learning, and deep learning, etc. What kind of problem is suitable to be solved with artificial intelligence technology? To put in simple terms, fuzzy theory mainly concerns with the deduction of possibility and uncertainty, the artificial neural network excels in classification and learning problems, the cluster analysis is good at classification, decision tree analysis is for classification and regression forecast, association rule learning is suitable for finding relationship between features, the advantage of genetic algorithm lies in optimized design, the expert system is suitable to reason and systematically solve problems of small scope while deep learning works for a large amount of data.

Data mining is the branch of artificial intelligence technology and the process to discover and to isolate interesting and valuable problems from big data. Based on analyzing a large amount of data to find meaningful relationships or rules using an automatic or semi-automatic way. Besides, there is another term knowledge discovery in databases (KDD) with similar meaning. The method of KDD is the process of explored useful knowledge from the database [1] and this method widely used data mining techniques.

B. Chi-square Test and Decision Tree

The Chi-square test was first proposed by British statistician Karl Pearson in 1900 [2]. The method was used to check relationships between categorical variables. It is suitable for category data, such as gender, party preferences, or religion.

The decision tree was an algorithm in machine learning. It was also a common data mining technique and usually used for classification and prediction. It was easy to implement, easy to explain and fit with human intuitive thinking. The decision tree method was first proposed by Quinlan. It was ID3 at first, and ID3 was the simplest algorithm of the decision tree [3]. Besides, the common algorithms for several decision trees included ID3, C4.5, C5.0, CART, and Random Forest [4].

The elements constituting the decision tree include nodes, internal nodes, and branch nodes. The nodes represent all the case samples, the internal nodes correspond to the feature attributes such as design attributes while the branch nodes represent the decision or result of classification and have choices just like the roads with forks. Hence, this kind of supervised learning algorithm base on if-then-else is similar

to formal expert systems. However, the rules for decision trees have been automatically generated by training computers and were not defined by experts or program designers.

C. Automotive Style

The different periods of cars will have different styles. For example, 1904 Mercedes and the 1904 Peerless Green Dragon were precursors of a change in form design, where cars became longer and lower, with long bonnets at the front with a relatively small passenger area behind [5]. During the 1930s and 1940s, vehicles become rounder and more integrated [6].

The design history records the streamlined style from 1935 to 1955. The design features of the streamlined style are filled with a form of the curve and long horizontal lines. The streamlined style pushes the consumer revolution in the 1950s and becomes a visual language of American modernity. Streamlining was presented as hopeful, a clear and proud statement about moving towards the future in the present day at that particular moment in time [7].

Regarding the design features, we take the Coupe Deville model of Chrysler in 1959 as an example to explain the design features of the rocket tail. The rocket tail of the vehicle covers the 1950s and 1960s, with the highest peak between 1955 and 1961. This was a style that considered the golden age of the development of American design.

The rocket tail has captured the imagination of the vehicle buyer, so auto companies attempted to design larger and larger rocket tail in the new models. With the advent of the jet aircraft and rocket, the design of the vehicle tail was more and more similar to the jet aircraft and rocket. This was an interesting situation. The vehicles with a rocket tail were just significantly decorated with a symbol of the spirit of the era.

Another thing to mention is the difference between similar design features. The rocket tail and the tailfin are both features at the tail of the vehicle, and they are different from the arc design of the streamlined style and the square design of the new design style. Then it depends on the judgment of the researcher of product design history to classify the rocket tail or tailfin.

British researcher Dowlen was used the vehicles design database to provide evidence of various style evolutions and innovations and described the evolution by time series. The vehicles were classified and applied cluster analysis using the CAPTCA method. This approach would find similarities of vehicles based on the classified clusters [5]. To discover historically innovative designs and found out the interesting vehicles [8], but did not describe the rules of vehicle style.

III. MATERIALS AND METHODS

Vehicles could be divided into different styles and eras: 1875 motor tricycles style, 1904 Ford model T style, 1935 streamlined style, 1955 popular style, 1975 new design style, and contemporary style. Computers could recognize different styles at different periods if we provide enough information. In the first stage, we extracted different design features of each vehicle, which was the engineering feature generally mentioned. In the second stage, the design features were

input into the computer program and then obtained the answers regarding whether the design features and style were related.

We take the historical style of vehicle design as an example. The study follows the historical context of automotive design. Applying the concept of a design database to collect cars by Dodge in the past and captured important features of the vehicle design. Based on the design source, some design features of appearance changed with design style.

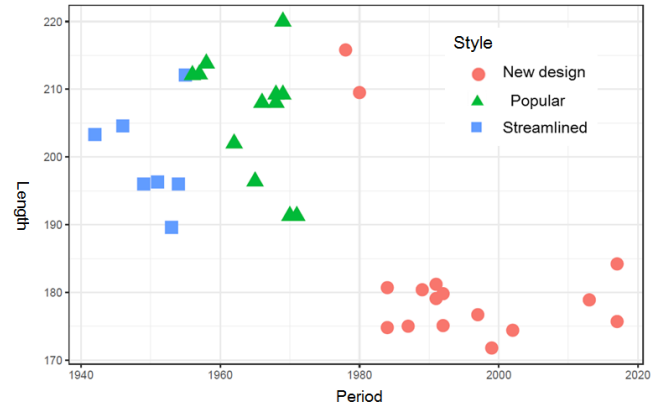


Fig. 1. Correlation of length and time series of Dodge 35 Cars between 1942 and 2017.

We could divide the vehicles into three historical styles, streamline style, popular style, and new design style. The samples selected are Dodge cars manufactured between 1942 and 2017. Although there are still many classic cars from the Dodge automobile company earlier, because the detailed information is not easy to collect, it is only collected the data of Dodge vehicles during the 70 years.

This investigation has three topics. The first, using the Chi-square test to discover design features affecting automotive style. The second, using a representative decision tree method to generate 50 decision trees, and choose a representative decision tree from them. The third, compare the design features that affect the style of the three methods, Chi-square test, one representative decision tree and calculate design features from 50 decision tree models.

IV. RESULT AND DISCUSSION

A. Chi-square Analysis to Discover Design Features Affecting Automotive Style

This investigation chose eight external features of the cars including vehicle length, fender design, number of headlamps, rear form, position of quarter glass, engine hood scoop, rocket tail, and side decoration. These eight external features were extracted from ten features mentioned in the article "Strule – a classification system for car styling" published in Journal of Design [9]. We looked for these design features from these Dodge car samples.

Two of the features include the length and number of headlamps are numerical while the other design features are categorical. After selecting the eight external features, we used Chi-square independence test as the verification for decision tree analysis and confirm whether a relationship is present between the design style and design feature.

TABLE I: APPLYING THE CHI-SQUARE TEST TO CHECK THE RELATIONSHIP BETWEEN DESIGN FEATURES AND STYLE

Characteristic	Length	Fender	Number of headlamps	Sideways decoration
p-value	0.000002	0.0001	0.008	0.02
Relationship	○	○	○	○
Characteristic	Back form	Position of quarter glass	Engine hood scoop	Rocket tail
p-value	0.00002	0.00002	0.14	0.004
Relationship	○	○	×	○

Table I shows the result of the Chi-square independence

test. Since the p-value is less than 0.05, the null hypothesis is rejected and hence there is significant evidence suggesting that the style and design features are related.

The analysis is the Chi-square test performed using R 3.6.1. The result shows that out of the eight appearance features, only the p-value of the engine hood scoop larger than 0.05. As a result, the style is not related to the engine hood scoop design but is related to other design features. If there is no significant correlation, show a cross symbol and there is a significant correlation, show a circle symbol in the table. It can use these design features as a comparison by the subsequent decision tree analysis.

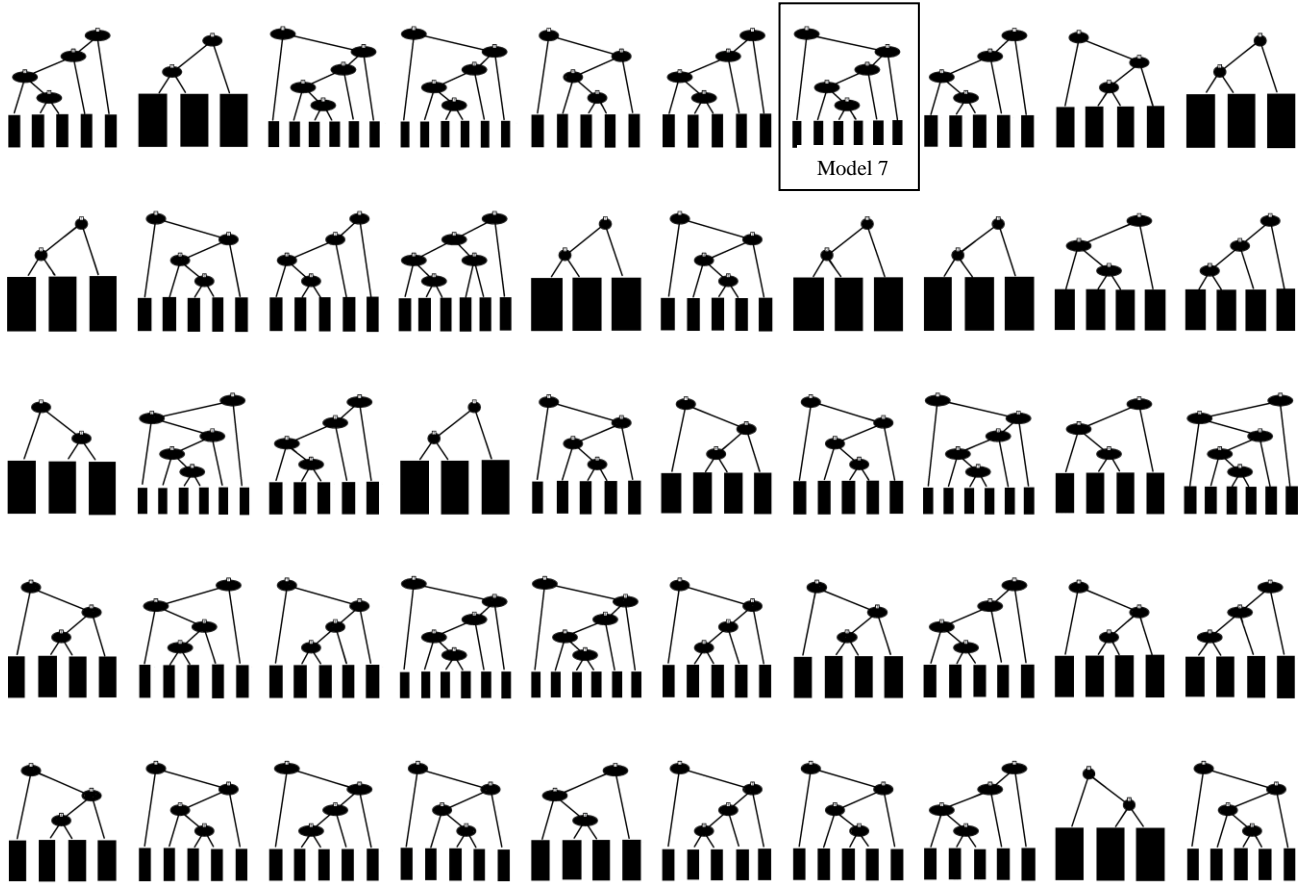


Fig. 2. 50 Decision trees generated automatically by the computer and black frame showed as model 7.

TABLE II: THE AVERAGE OF THE ACCURACY OF 50 DECISION TREES, THE NUMBER GRAY COLOR SHOWED AS MODEL 7

Training Models									
96	92	100	100	96	100	96	96	92	92
88	96	96	96	88	96	88	92	96	88
92	100	96	96	96	96	96	96	96	96
92	96	96	96	96	96	96	96	96	88
96	96	96	96	96	100	96	96	88	96
Average of accuracy rate is 94.96									

Testing Models									
80	50	90	90	90	70	100	80	70	60
80	80	70	100	70	80	60	70	70	90
50	80	60	50	80	60	100	70	90	70
80	80	90	100	80	100	80	70	70	70
60	80	80	90	40	70	90	60	80	90
Average of accuracy rate is 76									

B. Representative Decision Tree Method

Next, we will introduce a representative decision tree method, which is changed from C4.5 / C5.0. We do not use the design features of chi-square analysis as input attributes of the decision tree. Instead, we input all features into the computer and let it automatically generate design features.

Hence, we set the appearance features such as length (L1), fender (F1), number of headlamp (H1), back form (B1), position of quarter glass (Q1), rocket tail (R1), engine hood scoop (E1) and side decoration (S1) as classification attribute.

The classification of automotive-style includes streamlined style, popular style and new design style and they are used as the target attributes. The decision tree C5.0 package in R 3.6.1 is used to classify the 35 Dodge vehicles automatically and the result is as shown Fig. 2.

Table II shows the count for the average accuracy of the 50 samples each time. This includes the aforementioned eight

features and the decision tree shape, classification attribute and decision style are automatically generated as output. The style decided by the computer was compared with the style documented in design history to obtain accuracy.

Fig. 2 shows fifty decision trees generated automatically by the computer. We also would obtain knowledge of design from the data. The accuracy of the training mode ranges from 88% to 100% while the accuracy of the test model ranges from 50% to 100%. The statistics showed that the average training sample has an accuracy of 94.96% while the average test sample has an accuracy of 76%.

Based on the level distribution of the node in the decision tree diagram of the most representative decision tree (Model 7), one can know that the order of information gain for the decision tree is L1, F1, Q1, R1, and lastly B1. This means that L1 is the main factor determining the different automotive design style, followed by Q1. The third would be F1, fourth was R1, fifth was B1.

C. The Cumulative Frequency in the 50 Decision Trees

From the results above, we could get the Chi-square test and representative decision tree methods to find the design features that influence the car style. The representative decision tree method had two effects. The first effect was observing which design features of the decision tree were with high accuracy and the second effect was the calculated design features of 50 decision trees.

Based on the calculated design features of 50 decision trees. L1 had the highest cumulative frequency followed by F1, R1, B1, Q1, S1, and H1. One of the design features namely E1 was not considered as classification features in the computer-generated decision trees. Fig. 3 shows the cumulative frequency of features in 50 decision trees.

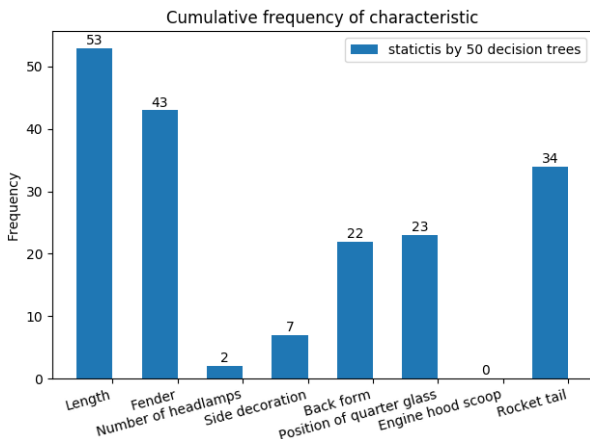


Fig. 3. The cumulative frequency of each design feature in 50 decision trees.

D. Comparison Decision Tree Method and Chi-square Test in Features that Influence Vehicle Style

Next, we compare the result with the chi-square result calculated earlier. It was found that the p-value is 0.15 which is larger than 0.05, suggesting that the design features from decision trees randomly generated by the computer are not the design features chi-square tests deemed related to style.

Finally, we can compare the results of the Chi-square test and representative decision tree methods. As each design features that affect the vehicle style will not be consistent using different methods. We found that the length was the

most important feature relative to the vehicle style from the three methods, chi-square test, representative decision tree or count the design features of 50 decision trees. But the engine hood scoop was the less important feature relative to the style among the three methods. Besides, fender design ranked second by calculated design features from 50 decision trees but ranked fourth by the chi-square test.

TABLE III. COMPARISON CHI-SQUARE TEST AND REPRESENTATIVE DECISION TREES IN DESIGN FEATURES AFFECTING AUTOMOTIVE-STYLE

Ranking	Chi-square test	Representative decision tree	Statistic of 50 decision trees
1	Length	Length	Length
2	Position of quarter glass	Position of quarter glass	Fender
3	Back form	Fender	Rocket tail
4	Fender	Back form	Position of quarter glass
5	Number of headlamps	Rocket tail	Back form
6	Rocket tail	Number of headlamps	Side decoration
7	Side decoration	Side decoration	Number of headlamps
8	Engine hood scoop	Engine hood scoop	Engine hood scoop

V. CASE DISCUSSION

We take the Dodge Challenger as a case study to discuss the vehicle sample using a representative decision tree method, Chi-square analysis method, and statistic of 50 decision trees method. The case study also obtains two summarizes, design features with less correlate to the vehicle style and uncertain design styles.

A. Dodge Challenger R/T

1) The method of chi-square test

First, we attempt to use the method of Chi-Square test to analyze the Dodge Challenger R/T. We can know the vehicle length, the position of quarter glass, back form and fender design are design features with a high correlation with historical style from the results of the chi-square test. The car for Dodge Challenger R/T belongs to modern cars with new design style due to the manufacture year 2010.

Such as a car with a long size, 197.7 inches in length, is more like popular style and streamlined style. Then some of the design features include the position of quarter glass at the rear, back form is square, and without fender design do not belong streamlined style. So we can predict the results from time series and Chi-square test that Dodge Challenger R/T is a modern vehicle but with significant characteristics of popular style.

2) The method of representative decision tree

Then we discuss this case with representative decision tree method methods. This Dodge Challenger R/T carcass was produced by Dodge in 2010. The product is characterized by a length of 197.7 inches, a headlight number of four, a quarter glass position at the rear, no fender design, and a square tail design. There are hood vents, no rocket-type tails, and decoration in the sideways, which is classified as a new

design style (possibility 2/3) or popular style (possibility 1/3). The judgment is correct.

But why is it that the car has two styles of features, the following is the reason for the case.

(1) The length of the Challenger R/T is 197.7 inches, which is close to the popular design of the car length in the popular period. (2) The Challenger R/T has an engine hood vents, a number of headlights of four, and the position of quarter glass in the rear. These features, along with the sideways decoration, feel a bit retro style.

3) The method of the statistic of 50 decision trees

Finally, we discuss the design features from the viewpoint of the statistic of 50 decision trees. The design features with highly-ranking include vehicle length, fender design, rocket tail design, and the position of quarter glass. However, the vehicle length and without fender design of Dodge Challenger R/T tend to classify the case as to be popular style. And the rocket tail is a unique design feature of the popular style. But Dodge Challenger R/T does not have this design feature, and it is difficult to let us find the correlation between this vehicle and popular style. Thus based on the three design features, it appears different opinions on vehicle style.

Then observed the next design feature as the position of quarter glass at the rear. Due to the clue of this feature, there is more evidence to confirm that its style is biased towards popular style. And the position of quarter glass at the rear is also a design feature ranked second by the chi-square test and the representative decision tree method. It can be said that it is an important feature used to determine its vehicle style.



Fig. 4. 2010 Dodge Challenger R/T car (From Wikipedia) [10].

B. Uncertain Style

As shown in Fig. 4, the decision of Dodge Challenger R/T is the uncertain style classified by the decision tree. This is because there are cars with the same features but belonged to a different style in the training samples. Two of the three samples belong to the new design style while one of them belongs to the popular style. As a result, when the test sample contains a feature of the same category, the computer will classify it as an uncertain style based on the result obtained in the training mode. In real life, one possible scenario is the transition between styles, producing a mixed style. The other possible scenario would be similar to a replica where the company made amendments to the once-popular product and allowed the product to return on the market. The gap could be several years and hence cross the two design style periods, producing such a result.

C. Design Features with Less Correlate to the Style

As for the engine hood scoop, although these three methods include chi-square analysis, representative decision tree method, and calculation of 50 decision tree methods. All

of these methods did not think engine hood scoop a significant design feature relevant to the style. But Dodge Challenger R/T, engine hood scoop plays an important and significant design feature in this case.

This feature appeared in the later period of popular style and had a function to lower the engine temperature. However, its appearance is relatively short, and it is not common in new design style cars. Thus we can explain why the correlation is not significant. Because of all the Dodge car samples, there are fewer vehicles with such a design feature. Therefore, the engine hood scoop is not considered to be a design feature with significant style relevance. But in fact, this design feature is very special.

VI. APPLICATION

Where can this study of design knowledge be applied to? By applying the three methods mentioned above, we could learn the following:

- 1) We knew that the design features are related to car style by using Chi-square test, based on p-value, it was possible to obtain the ranking of design features and its importance.
- 2) We could also obtain the design features of the decision tree with the highest accuracy by the representative decision tree method.
- 3) We can also obtain the cumulative frequency of design features from 50 decision trees statistics.
- 4) We could predict the style of design and design new vehicle appearance using design knowledge.

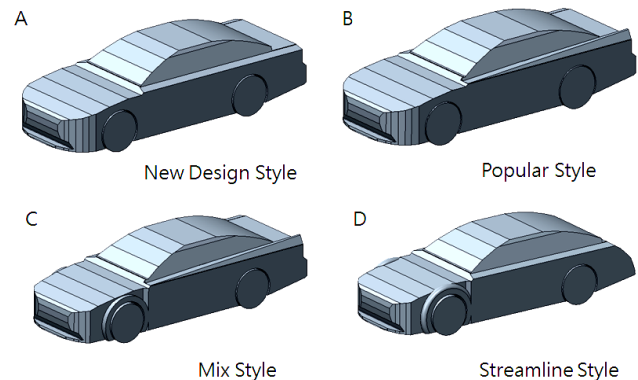


Fig. 5. Applying knowledge of automotive style to design a new car appearance.

Fig. 5A is a concept of a modern car with a new design style, and Fig. 5B is added with a rocket tail design, giving the model one feature of the popular style. Fig. 5C is a combination of a rocket tail and a fender, giving the model a mix of two styles. Fig. 5D is an arc design of the rear design, giving it a feeling of streamlined style.

VII. CONCLUSIONS

A clear style classification principle is important and useful for designers. As a result, this investigation attempts to incorporate Chi-Square statistic method, the method of representative decision tree analysis, and the method of the statistic of 50 decision trees in design research with design styles of Dodge automobiles as the case study. To understand

the rules behind the styles, certain design characteristics can be identified. This will help us to understand the design style and identify the retro or mixed style. After comparing the styles from the history of design, we can try to use machine learning to obtain design style classification. This can be likened to using a rational view to observe an emotional problem.

This investigation also found the design features that influenced car style. It would be possible to generate a new automotive-style or mixed style by selecting the important features. Thus we could keep our brand style and distinguish another competitor's style while use the method to classify, control and to understand the design styles. It is worthwhile to further discuss how this similar approach can be extended to other products.

AUTHOR DECLARATION

The author declares another conference paper "rule induction of automotive historic styles using decision tree classifier" under review and unpublished in conference ICCCI 2020, which discusses the theoretical part of automotive historic styles using decision tree classifiers. The content is about the rule induction of the decision tree and using the confusion matrix to evaluate the model. This article is about the comparison of design features influencing vehicle style using decision tree method and chi-square test method to elaborate on how to apply design knowledge to automotive conceptual design. These two papers use the same vehicle database.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

The authors built the framework of this research together. Dr. Wang planned, organized and supervised the content of the paper, and Chen was responsible for collecting automotive-related data to perform statistics and operation of machine learning using R and Python. All authors had approved the final version.

ACKNOWLEDGMENT

The author is extremely grateful for the technical

guidance on the machine learning from Stella Lin in NTU computer and information networking center.

REFERENCE

- [1] U. Fayyad, G. P. Shapiro, and P. Smyth, *From Data Mining to Knowledge Discovery in Databases*, 1996.
- [2] J.-Y. Dai, *Statistical Concepts and Methods*, 1992.
- [3] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.
- [4] L. Rokach and O. Maimon, *Decision Trees*, Tel-Aviv University, 2005.
- [5] C. Dowlen, "Creativity in car design – the behaviour at the edges," presented at the The 2nd International Conference on Design Creativity (ICDC2012), Glasgow, UK, 2012.
- [6] C. Dowlen and J. Shackleton, "Design history of the car: an empirical overview of the development of layout and form," in *Proc. International Conference on Engineering Design*, Stockholm, 2003.
- [7] W. Kowalik, "Streamline moderne design in consumer culture and transportation infrastructure: design for the twentieth century," 2017.
- [8] C. Dowlen, "Design paradigms in car history," PhD thesis, London South Bank University, 2017.
- [9] H.-H. Wang and C.-P. Chen, "A classification system of car style," *Journal of Design*, vol. 9, no. 2, 2004.
- [10] Wikipedia page regarding Dodge Challenger. [Online]. Available: <https://en.wikipedia.org/>

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Hung-Hsiang Wang is an associate professor in the Department of Industrial Design at the National Taipei University of Technology, and the director of Product Design and Development Laboratory. His research interests are cultural and creative product design, design creativity and design thinking.



Chip-Ping Chen is a Ph.D. student in the College of Design at National Taipei University of Technology. His major is industrial design and currently work on research projects related to knowledge discovery in databases, pen-based computer interaction, and product mechanical design.