ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

# Research on Optimization of AI Image Recognition Performance Based on Multiple Machine Learning Algorithms and Deep Learning Models

**Zhuoyang Li**[1,*]

[1]*School of Statistics and Data Science, Qufu Normal University, Jining, China*
[*]*Corresponding author: 2067598048@qq.com*

***Abstract:*** *This paper explores the growing use of AI-generated images for fraud and the spread of piracy, while AI generated images are widely used in various fields. By using principal component analysis (PCA) to reduce feature dimension, shallow learning methods such as Logistics regression, discriminant analysis, SVM (support vector machine), random forest, KNN (K-nearest neighbor), and deep learning methods with attention mechanism such as Alexnet, Googlenet and Mobilenet are adopted. The AI-generated picture is effectively distinguished from the real picture. The results show that SVM model and Alexnet model show the strongest comprehensive performance in identifying AI-generated images, providing new ideas and methods for solving problems caused by AI-generated images.*

***Keywords:*** *Principal component analysis, deep learning, Alexnet, Mobilenet, Pros and Cons Solution distance method (TOPSIS)*

## 1. Introduction

In recent years, with the rapid development of AI technology, AI-generated pictures are increasingly widely used in various fields. In order to cope with this challenge, scholars and scientific research institutions at home and abroad have invested in research, aiming to develop methods and technologies that can effectively distinguish AI-generated pictures. In this context, based on machine learning algorithms, this paper comprehensively evaluates the performance of a variety of models in recognizing AI-generated images. In this paper, the Reset model is first used to convert image data into CSV data, and principal component analysis (PCA) is used to reduce the dimension of feature vectors [1]. Then, shallow learning methods such as Logistics regression [2], discriminant analysis, SVM (support vector machine), random forest [3], KNN (K-nearest neighbor) [4] and deep learning methods with attention mechanism such as Alexnet [5], Googlenet [6] and Mobilenet [7] are adopted. By means of cross-validation and Bayesian optimization, the classification efficiency and classification ability of these models are studied and compared in depth.

Through the research in this paper, we expect to provide new ideas and methods for the identification of AI-generated pictures. To a certain extent, promote the development of machine learning algorithm in the field of AI-generated image recognition, deepen the understanding of machine learning on AI-generated images, and promote the further practice and development of machine learning methods in related fields.

## 2. Model framework

### *2.1 Shallow layer method*

(1) Logistic regression

Logistic regression, also known as logistic regression or log-probability regression, is a classification algorithm commonly used in statistical learning and machine learning, primarily for solving binary classification problems. The basic idea of Logistic regression is to use the available data to establish regression formulas for classification boundaries in order to classify. Specifically, it predicts the relationship between a dependent variable (usually a binary class) and an independent variable (i.e. a feature) by fitting a logical function (usually a Sigmoid function). A logical function can map any real number between 0 and 1, thus representing the probability of an event occurring. The formula for the

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

Logistic regression model to solve the binary classification problem is arranged as follows.

$$\begin{cases} \hat{p} = \sigma\left(\theta^T x\right) = \dfrac{1}{1+e^{-\theta^T x}} \\ \hat{y} = \begin{cases} 1, \hat{p} \geqslant 0.5 \\ 0, \hat{p} < 0.5 \end{cases} \end{cases} \tag{1}$$

The final criterion is shown in the following formula.

$$\hat{y} = \begin{cases} 1, \theta^T x \geqslant 0 \\ 0, \theta^T x < 0 \end{cases} \tag{2}$$

(2) Secondary discriminant analysis

Quadratic Discriminant Analysis (QDA) is a statistical method and machine learning algorithm that is primarily used to solve classification problems. It is built based on the assumption that the sample data of the class obey a multivariate normal distribution, and differs from linear discriminant Analysis (LDA). An important feature of QDA is that it does not assume that the classes have the same covariance matrix. Instead, QDA computes a definite covariance matrix for each category and brings it into the discriminant function for classification. Because QDA takes into account the differences in the covariance matrix within the class, it can provide better classification results than LDA in some cases. The steps of the secondary discriminant analysis are as follows.

$$\mathrm{P}(y=k) = \frac{\sum_{c=1}^{n} 1\{y^{(c)} = k\}}{n} \tag{3}$$

$$\mu_k = \frac{\sum_{c=1}^{n} 1\{y^{(c)} = k\} x^{(c)}}{\sum_{c=1}^{n} 1\{y^{(c)} = k\}} \tag{4}$$

$$\sum_k = \frac{\sum_{c=1}^{n} 1\{y^{(c)} = k\}(x^{(c)} - \mu_{y^{(c)}=k})(x^{(c)} - \mu_{y^{(c)}=k})^T}{\sum_{c=1}^{n} 1\{y^{(c)} = k\}} \mathrm{P}(y = k \mid X = x) \tag{5}$$

$$\delta_k(x) = log(\pi_k) - \frac{1}{2}\mu_k{}^T \sum_k{}^{-1} \mu_k + x^T \sum_k{}^{-1} \mu_k{}^{-\frac{1}{2}} x^T \sum_k{}^{-1} x - \frac{1}{2} log \mid \sum_k \mid \tag{6}$$

Quadratic discriminant analysis can use the above formula to estimate the corresponding probability. The expression $\mathrm{P}(y = k \mid x)$ can not be divided because of the difference $\sum$. The discriminant equation of quadratic discriminant analysis is quadratic equation about $x$, and the classification boundary is curve.

(3) SVM (Support Vector Machine)

SVM, full name Support Vector Machine (support vector machine), is a generalized linear classifier that classifies data according to supervised learning (supervised learning). Its decision boundary is the maximum-margin hyperplane on which the learning sample is solved.

The basic steps of SVM (support vector Machine) are as follows:

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

$$\begin{cases} \text{Objective\_function}: \min\limits_{w,b,\xi} \dfrac{1}{2}\| w \|^2 + C\sum\limits_{i=1}^{l}\xi_i \\ s.t.\ \ y_i\left(\left(w\ \Phi(x_i)\right)+b\right) \geqslant 1-\xi_i, i=1,\cdots,l \\ \xi_i \geqslant 0, i=1,\cdots,l \\ \text{Decision\_function}: f\left(x\right)=sgn\left(\left(w^*\Phi(x)\right)+b^*\right) \end{cases} \tag{7}$$

In the formula (7), $\xi=(\xi_1,\xi_2,\cdots,\xi_i)^T, C>0$ is a penalty function, the objective function requires the support vector machine to be both minimized $\| w \|^2$ and minimized $\sum_{i=1}^{l}\xi_i$, and the parameters are used to balance these two requirements.

(4) KNN (K- nearest neighbor)

KNN (K-Nearest Neighbors) is an instance-based learning algorithm that can be used for both classification and regression. Its basic idea is to find the nearest $K$ samples by calculating the distance between a new sample and all the samples in the training set, and then determine the category or value of the new sample according to the category or value of this $K$ sample.

The following describes the calculation steps of the KNN algorithm by taking the classification sample $x_m$ as an example: Suppose that the data set of the training sample is $U=\{(x_i,c_i)\,|\,i=1,\ldots,N\}$, and suppose that the P-dimension column vector in the training set sample $\mathrm{C}=\{\mathrm{c_j}\,|\,\mathrm{j}=1,\ldots,\mathrm{M}\}$ is the corresponding $c_{\mathrm{i}}$ category label . The process of the KNN algorithm is briefly described as follows.

$$\mathrm{dis}(x_m-x_i)=\sqrt{(x_m-x_i)^{\mathrm{T}}(x_m-x_i)} \tag{8}$$

Secondly, determine the nearest neighbor parameter $k$, find the nearest neighbor $k$ from the sample to be classified, and assume the set $\mathrm{D_m}=\{(x_{mt},c_{\mathrm{mt}})\,|\,\mathrm{t}=1,\ldots,\mathrm{k}\}$, where the $t(1\leq t\leq k)$ nearest neighbor of the sample $x_m$ to be classified $c_{\mathrm{mt}}$ is the corresponding class $x_{mt}$.

Finally, voting is carried out according to the category of the nearest sample calculated in the set. Assuming the voting result is, the final voting result $v=[v_1,\ldots,v_{\mathrm{i}},\ldots,v_{\mathrm{m}}]$ is made in accordance with the principle of minority obedience to majority. The decision-making rule is as follows:

$$c(x_m)=\underset{i}{argmax}(\boldsymbol{v}_i) \tag{9}$$

(5) Random Forest

Random forest is an ensemble learning method based on decision trees that combines the predictions of multiple decision trees to improve the stability and accuracy of the model.

In the stage of model prediction for the test set, the classification accuracy rate is set as a, and the calculation formula is as follows.

$$\alpha=\frac{\sum_{\mu=1}^{\lambda}c_{\mu}^{'}\odot c_{\mu}}{\lambda} \tag{10}$$

***2.2 Deep Method***

(1) AlexNet

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

The AlexNet neural network model is a typical convolutional neural network model, which was first proposed by Alex Krizhevsky at the ILSVRC competition in 2012. AlexNet consists of 8 network layers, of which there are 5 convolutional and 2 fully connected hidden layers, and 1 fully connected output layer. AlexNet changed the sigmoid activation function to a ReLU activation function to make calculations simpler and the network easier to train, and controls the model complexity of the full connection layer by using dropout.

(2) GoogleNet

GoogleNet convolutional neural networks are a completely new deep learning structure proposed by Christian in 2014. The total number of layers used to build the network is about 100, the depth of the network is 22, and its main structure is composed of 9 association modules (inception model (IM)).

(3) MobileNet_V2

MobileNetV2 network is a lightweight convolutional neural network, with fewer parameters, fast reasoning speed, and can still achieve high segmentation accuracy under the condition of small data sets.

## 3. Data analysis

The data in this paper comes from the kaggle platform. In order to make the data can be used in shallow learning methods to train models and forecast, the picture data is preprocessed, and the data form is converted from picture data form to CSV data form. After that, the data were normalized.

For the transformed data, the dimensionality of the feature vector is up to 1000, so principal component analysis (PCA) is used to reduce the dimensionality of the data. We classify classification methods in machine learning into shallow learning methods and deep learning methods based on whether they can directly process image data. Shallow learning methods are classification methods that cannot directly process image data, while deep learning methods are classification methods that can directly process image data.

### *3.1 The use of shallow learning methods*

In order to select the method with the best classification ability and the highest classification efficiency in terms of identifying AI-generated images among the shallow learning methods, a classification learner is used to classify the data after dimensionality reduction of principal component analysis using various machine learning classification methods. Since cross-validation divides the data into training sets and test sets through multiple iterations, the performance of the model can be evaluated more comprehensively. And each sample will be used as a part of the training set and test set, maximizing the use of the information in the data set, can better evaluate the performance of the model on the unseen data, that is, the generalization ability of the model. Therefore, we adopt the method of cross-validation, set the number of cross-validation to 10, and adjust the parameters of the model after running, observe the difference between the classification ability and classification efficiency of the model before and after adjustment, and select the best machine learning algorithm in the field of identifying AI-generated images.

*Table 1: Classification accuracy and time of different learning methods*

| Classification Method | Classification accuracy | AUC value | Classification time (in s) |
|---|---|---|---|
| SVM | 88.4% | 0.937 | 65.63 |
| SVM (after Bayesian optimization) | 88.6% | 0.936 | 8009.10 |
| KNN | 85.3% | 0.897 | 9.44 |
| KNN (after Bayesian optimization) | 85.7% | 0.914 | 290.56 |
| Quadratic discriminant analysis | 83.0% | 0.906 | 4.10 |
| Discriminant analysis (after Bayesian optimization) | 83.0% | 0.906 | 46.59 |
| Random forest | 81.9% | 0.865 | 143.71 |

In summary, compared with before Bayesian optimization, the AUC values of the three models are

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

not significantly different. The time of the three models in the process of Bayesian optimization is longer than the previous classification time. In terms of classification accuracy, after Bayesian optimization, the classification accuracy of SVM (support vector machine) model and KNN model is improved, while the classification accuracy of discriminant analysis model remains unchanged.

In order to further observe the experimental results, the optimal learning method in identifying AI-generated images is selected, and the specific values are listed in Table 1. As can be seen from Table 1, the SVM model after Bayesian optimization and tuning has the highest classification accuracy, but also the longest time, the AUC value is close to the maximum, and the quadratic discriminant analysis takes the shortest time. Based on the above experimental results, among the above shallow learning methods, SVM (support vector machine) model after Bayesian optimization tuning has the best prediction performance, and the classification efficiency of quadratic discriminant analysis is the highest, but the classification accuracy is lower than that of SVM model after Bayesian optimization.

### *3.2 Use of deep learning methods*

Because the data set is too large, the training time of deep learning method is too long, so we conducted hierarchical sampling among 80,000 pictures in the training set and 20,000 pictures in the test set to extract the data representing the test set and the training set. Among them, 10,000 pictures were sampled from each of the two kinds of pictures in the training set as the new training set, and 1100 pictures were sampled from each of the two kinds of pictures in the test set as the new test set for the training and testing of the model. In the process of model training, the input dimension of the control data was (mini-batch,3,32,32), the training epoch of the model was 10, the learning rate was 0.0001, the mini-batch was 32, and Alexnet with attention mechanism was used at the same time. Googlenet and Mobilenet models were trained. We saved the models with the highest test set AUC value among several models and visualized them in ROC matrix through the confusion matrix of these models. The three deep learning models were evaluated according to the classification accuracy, classification time and AUC value of the models.

Since the AUC value and classification accuracy of a model are both important indicators used to evaluate the performance of a classification learner, here we discuss the classification performance of the model with the AUC value. As can be seen from Table 2, among the three models, Mobilenet has the highest classification performance, with an AUC value of 0.9649, followed by Googlenet and Alexnet, with an AUC value of 0.9609 and 0.9594, respectively.

From the perspective of classification time of the models, the classification time of Alexnet with the highest classification efficiency is 9500 seconds, followed by Googlenet and Mobilenet with 21536 seconds and 25760 seconds, respectively.

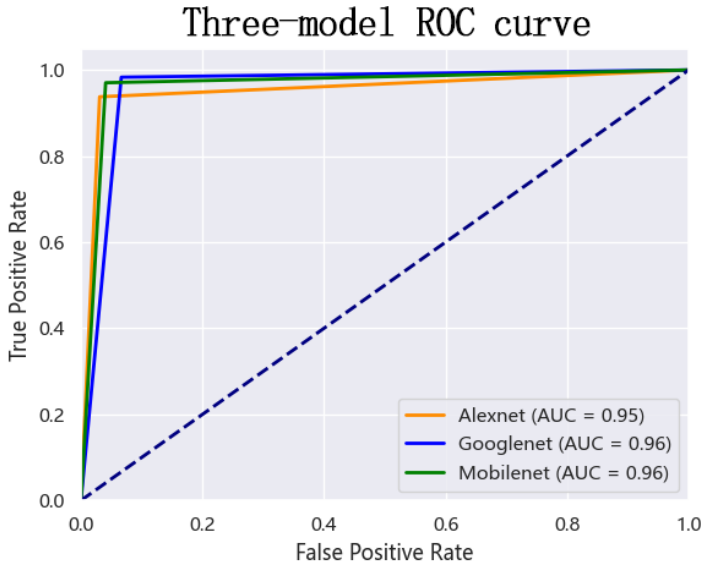


*Figure 1: Neural network structure*

As shown in Figure 1, the difference between the AUC values of the test sets of the three deep learning models is very small, indicating that there is almost no difference in the classification ability of the

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

models in hand. We need to determine the performance of the model from the classification time of the model, and the classification time of Alexnet model is only 9500 seconds, which is more than two times more efficient than the Googlenet model, which is the second most efficient. Therefore, the overall performance of Alexnet model is the best among the three deep learning models.

*Table 2: Classification performance and classification time of deep learning models*

| Data Dimension | Classification method | Classification accuracy | AUC value | Sorting time (s) |
|---|---|---|---|---|
| (2000,32,32) | Alexnet | 95.94% | 0.9594 | 9500 |
| | Googlenet | 96.09% | 0.9609 | 21536 |
| | Mobilenet | 96.49% | 0.9649 | 25760 |

From the classification results of AI-generated images and real images processed by the above machine learning methods in the big data environment, it can be seen that the classification accuracy of the shallow learning method is significantly lower than that of the AUC value and the deep learning method, and the classification efficiency of the deep learning method is lower than that of the shallow learning method.

## 4. Comprehensive evaluation of each classification method based on entropy weight method and TOPSIS

### *4.1 Entropy weight method*

The basic principle of entropy weight method is based on the concept of information entropy [8]. Information entropy is an index to measure the degree of information concentration. It represents the maximum degree of information in a random variable when the sum of probabilities of each value is 1. The smaller the information entropy is, the more concentrated the information is, and the other way around, the more dispersed the information is. The entropy weight method calculates the information entropy of each index, and then determines the weight of each index, so as to achieve multi-index decision-making.

First, 5 kinds of shallow analysis methods, the shallow analysis method after Bayesian optimization adjustment and 3 kinds of deep learning methods are identified as the evaluation object, the classification time and AUC value are identified as the evaluation index for forward processing and the forward matrix X is constructed. Secondly, since the classification time and AUC value are both positive numbers, Therefore, the forward matrix X is substituted into the following formula for standardization processing to obtain the standardized matrix Z.

$$z_{ij} = x_{ij} \Big/ \sqrt{\sum_{i=1}^{n} x_{ij}^2} \tag{11}$$

Subsequently, we calculate the probability matrix Z by substituting each element of the standardized matrix P into the formula.

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}} \tag{12}$$

Finally, we plug each element of the probability matrix P into the first two formulas to calculate the information entropy, information utility value. The weight of each index can be calculated by substituting the last formula for normalization, as shown in Table 3.

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln(p_{ij}) (j = 1, 2) \tag{13}$$

$$d_j = 1 - e_j \tag{14}$$

$$\omega_j = d_j / \sum_{j=1}^{m} d_j (j = 1, 2) \tag{15}$$

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

*Table 3: Weight distribution diagram of entropy weight method*

| Evaluation index | Information entropy | Information utility value | Weight |
|---|---|---|---|
| AUC values | 0.929 | 0.071 | 51.51% |
| Sorting time | 0.933 | 0.067 | 48.49% |

### *4.2 The advantages and disadvantages of distance solution (TOPSIS)*

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), also known as the method of sorting by similarity to ideal solution, is a commonly used multi-attribute decision analysis method. The basic principle of TOPSIS method is to sort the evaluation object by detecting the distance between the optimal solution and the worst solution.

For the normalized matrix $Z$ obtained above, we define the maximum and minimum values as $Z^+$ and $Z^-$ respectively.

$$Z^+ = \left(\max\{z_{11}, z_{21}, \cdots, z_{11\,1}\}, \max\{z_{12}, z_{22}, \cdots, z_{11\,2}\}\right) \tag{16}$$

$$Z^- = \left(\min\{z_{11}, z_{21}, \cdots, z_{11\,1}\}, \min\{z_{12}, z_{22}, \cdots, z_{11\,2}\}\right) \tag{17}$$

Then, the maximum $Z^+$ and minimum $Z^-$ values obtained and the weight of each index obtained $\omega_j$ by entropy weight method are substituted into the following formula to calculate the score of an evaluation object $i$, where $i = 1, 2..., 11$.

$$D_i^+ = \sqrt{\sum_{j=1}^{2}\left(Z_j^+ - z_{ij}\right)^2} \tag{18}$$

$$D_i^- = \sqrt{\sum_{j=1}^{2}\omega_j (Z_j^- - z_{ij})^2} \tag{19}$$

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{20}$$

*Table 4: Ranking of comprehensive evaluation scores of each classification method*

| Classification Method | Positive ideal Solution Distance (D+) | Negative ideal solution distance (D-) | Composite score index | Sort |
|---|---|---|---|---|
| SVM | 0.20 | 0.87 | 0.81 | 1 |
| SVM (after Bayesian optimization) | 0.30 | 0.7 | 0.70 | 3 |
| KNN | 0.49 | 0.73 | 0.60 | 8 |
| KNN (after Bayesian optimization) | 0.37 | 0.77 | 0.68 | 4 |
| Quadratic discriminant analysis | 0.42 | 0.76 | 0.64 | 6 |
| Discriminant analysis (after Bayesian optimization) | 0.42 | 0.76 | 0.64 | 7 |
| Random forest | 0.71 | 0.69 | 0.49 | 11 |
| Binary Logistic Regression | 0.40 | 0.76 | 0.65 | 5 |
| Alexnet | 0.26 | 0.81 | 0.76 | 2 |
| Googlenet | 0.58 | 0.70 | 0.55 | 9 |
| Mobilenet | 0.70 | 0.72 | 0.51 | 10 |

The final comprehensive evaluation score results are shown in Table 4.

ISSN 2616-5775 Vol. 7, Issue 12: 51-58, DOI: 10.25236/AJCIS.2024.071207

From the above table, we can see that after evaluating the above machine learning algorithm model in recognition of AI-generated images from two aspects of AUC value (classification accuracy) and classification time, the SVM (support vector machine) model scores 0.81 among shallow learning methods, ranking first. Among all the above machine learning algorithms, it is the algorithm with the best comprehensive classification efficiency and classification time in identifying AI-generated images, followed by Alexnet model in deep learning methods, with a score of 0.76. Although the classification accuracy of deep learning method is nearly 100%, its huge classification time is a big problem in recognizing AI-generated images. Random forest algorithm is not dominant in the classification accuracy and classification time of identifying AI-generated pictures, and ranks 11th in the above algorithm model with a score of 0.49.

In summary, SVM (Support vector Machine) model and Alexnet are the two models with the strongest performance in identifying AI-generated images and real images among the above machine learning methods.

## 5. Conclusions

After discussing the classification accuracy, AUC value, classification time and the scores of entropy weight method and good distance method, this paper draws the following conclusions: SVM (support vector Machine) model and Alexnet model have the strongest comprehensive performance in identifying AI-generated images. The entropy weight method of SVM model and the distance method of Alexnet model rank first and second respectively. In any case, these two models are good choices for identifying AI-generated images. If the accuracy of AI-generated image recognition is pursued in the actual situation, the Alexnet model can be used because of its high classification accuracy; If you want to identify the speed of AI image generation, you can use SVM model because of its short training time.

## References

*[1] Aisha F, Yannis P, Michail T, et al. Cauchy robust principal component analysis with applications to high-dimensional data sets[J].Statistics and Computing, 2023, (1)*
*[2] Fan S, Jianqian C, Pei L, et al. Prognostic models for breast cancer: based on logistics regression and Hybrid Bayesian Network.[J].BMC medical informatics and decision Making, 2023, 23 (1) : 120-127.*
*[3] Xinmin M, Zhenyu C, Pan C, et al.Predicting the utilization factor of blasthole in rock roadways by random forest[J].Underground Space, 32, 2023112-245.*
*[4] Stelian N, Oana Sorina C, Lacramioara S .Recovery Gestures Classification Using KNN and LDA Models.[J].Studies in health technology and informatics, 2023, 30958-62.*
*[5] Sridharan V N, Vaithiyanathan S, Aghaei M .Voting based ensemble for detecting visual faults in photovoltaic modules using AlexNet features[J].Energy Reports, 2024, 113889-3901.*
*[6] Zhang K, Sun W, Ba Y, et al.Transformer Fault Diagnosis Method Based on SCA-VMD and Improved GoogLeNet[J].Applied Sciences, 2019, 14(2)*
*[7] Long W, Ming Z, Guangyuan H, et al. Classification of Breast Lesions on DCE-MRI Data Using a Fine-Tuned MobileNet [J].Diagnostics, 2019, 13(6):1067-1076.*
*[8] Hamdouni S, Benaicha M, Alaoui H A .Optimizing self-compacting concrete: formulation approach enhanced by entropy method[J].Discover Civil Engineering, 2019, 1(1):63-69.*