Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

Airfoil Aerodynamic Coefficient Prediction based on Ensemble Learning

Xingchen Yan [1, *] and Yuange Ma [2]
[1] College of Information and Intelligence, Hunan Agricultural University, Changsha 410000, China
[2] Department of Automation, North China Electric Power University, Baoding 071000, China
[*]Corresponding Author.

Abstract:

The calculation of wing aerodynamic coefficient is the main content of airfoil design and research, which is of great significance to improve flight performance. The traditional methods to obtain the aerodynamic coefficients of the airfoil by computational fluid dynamics method or wind tunnel test have the disadvantages of large calculation and high test cost. In recent years, the high-speed development of machine learning has proved that it has strong nonlinear mapping ability. Therefore, more and more scholars apply it to the prediction of wing aerodynamic coefficients. The ensemble learning algorithm in machine learning has a strong ability of classification, regression and generalization ability. Taking this into account, Random Forest (RF) and Extreme Gradient Boosting (Xgboost), which are cutting-edge in ensemble learning, are applied to the prediction of wing aerodynamic coefficients for the first time. Xgboost has higher promotion potential than RF, so this paper additionally adjusts the parameters of Xgboost and obtains the optimal training parameters. Finally, we compare the prediction accuracy between non-ensemble and ensemble learning algorithms. The experimental results show that the ensemble learning algorithms have higher prediction accuracy than the classical regression algorithms. Among them, the best algorithm is Xgboost, and the prediction accuracy of RF is slightly lower than Xgboost. The MAE, MSE, and RMSE of RF and Xgboost are approximately 10 ~ 100 times lower than that of other algorithms. In addition, Xgboost has lower time complexity and higher generalization capability.

Keywords*: Aerodynamic Coefficient; Random Forest; Extreme Gradient Boosting; Ensemble Learning*

I. INTRODUCTION

Airfoil aerodynamic coefficient calculation is the main content of airfoil design and research, which is of great significance to improve flight performance. The traditional methods for calculating aerodynamic coefficients of airfoils through Computational Fluid Dynamics (CFD) calculation or wind tunnel test is proved to be effective, but it has disadvantages of large computation and high test cost. With the rapid development of the neural network in recent years, its outstanding nonlinear mapping ability attracts more and more scholars to apply this method to the prediction of aerodynamic parameters. These include training wing parameters and training wing shape.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

A large number of scholars use wing parameters as feature inputs to predict wing parameters. In 2003, Suresh[1], based on recursive neural network modelling, predicted the lift coefficient of the rotor at a high Angle of attack and compared it with experimental data to prove the feasibility of his method. In 2011, Carpenter M's team[2] proposed a single hidden layer neural network for missile aerodynamic parameter prediction. Liu Xin[3] proposed a model based on RBF neural network and successfully applied it to the prediction of wing lift resistance in the vibration of the wing. Based on the optimized BP neural network, Yuan Zhijie et al[4] predicted the aerodynamic parameters of the missile and proved that the method has good generalization and fitting ability. Balla Kensley[5] proposed a multi-output neural network to predict the aerodynamic coefficients of 2D and 3D wings and compared it with the POD method. The results show that the neural network has better performance, especially in predicting the flow field containing shock waves.

In addition, there are also a large number of studies that use image data set and train CNN for classification. Zhang et al[6] used the convolutional neural network (CNN) to learn the lift coefficients of different airfoils at different angles of attack (AoA), Mach number and Reynolds number. Sekar et al[7] used CNN to approximate the flow field on the airfoil as a function of airfoil geometry, Reynolds number and AoA, without directly solving Navier Stokes equations. H. Chen et al[8] used a composite airfoil image data set generated by convolution of flow conditions to predict aerodynamic coefficients. Hui et al[9] used CNN to predict pressure distribution around an airfoil, while Guo et al[10] used CNN to predict non-uniform stable laminar flow in the 2D or 3D domain.

However, the data dimension and prediction accuracy of a single machine learning model is limited. In order to solve this problem, we consider the strong optimization performance of the ensemble learning algorithm and apply it to the prediction of airfoil parameters. Aiming to more objectively evaluate the better regression prediction ability of ensemble learning in wing aerodynamic parameters, this paper also compares the prediction accuracy of other classical non-ensemble regression algorithms and some ensemble learning algorithms in MSE, MAE and RMSE. At the same time, we also evaluated the time complexity of each model training, that is, the actual time of model training. We introduced the concept of time effect ratio(PTR), to more intuitively show the performance of each model on both training time cost and prediction ability. Ultimately, the experimental results show that the airfoil parameter prediction based on XGBoost has higher generalization potential and prediction accuracy than other ensemble learning algorithms and most other machine learning algorithms.

The article is mainly divided into four parts. The first part briefly introduces the background and application of machine learning and deep learning in airfoil parameters; In the second part, around the work done in this paper, we will introduce the ideas of two ensemble learning algorithms, random forest (RF) and extreme gradient boosting (XGBoost); The third part will show the experimental results of our work, including data set displaying, data analysis, parameter tuning and model comparison; For the fourth part, this paper will make conclusions, briefly describing the application prospect of machine learning and deep learning algorithms in this field.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

II. METHODOLOGY

2.1 Ensemble Learning

The core idea of traditional machine learning methods is usually to find the partition (super) plane of the data set or the mapping function between input features and output values. Nevertheless, in the practical application of machine learning algorithms, the model trained by a single learner under different initial parameter settings may not be optimal.

Consequently, to further improve the generalization performance and accuracy, Dasarathy and Sheela[11] first proposed the idea of ensemble learning in 1979. In 1997, Schapire and Freund[12] proposed a new boosting ensembling method. His boosting algorithm does not require any prior knowledge about the performance of the weak learning algorithm. It can overcome the low accuracy of each classifier singly and greatly enhance the accuracy of each classifier in a specific way. Since then, the research of ensemble learning has developed rapidly, and many novel ideas and models have emerged. Until 2001, Breiman[13] proposed the random forest (RF) algorithm, which classifies and regresses by integrating multiple random trees in parallel. Then, aimed at overcoming the limitations brought by the randomness of random forest, in 2016, Chen et al.[14] proposed a method of serial integration CART tree. When constructing the current tree, this algorithm introduces the prior knowledge of the previously constructed tree, so as to achieve better classification or regression accuracy.

In this paper, considering the superior regression performance of random forest and XGBoost, we apply them to the prediction of airfoil parameters.

2.2 Random Forest(RF)

In 2001, Breiman[15] first proposed the concept of random forest, whose basic modelling idea is to build different sample training sets based on the decision tree algorithm and generate a series of different decision tree models by combining randomly generated feature space. The biggest difference between random forest and other models is that it can establish multiple prediction models, so as to effectively avoid the over-fitting phenomenon of models and improve the performance of models. RF[16] is an ensemble learning algorithm further optimized on the basis of Bagging ensemble learning and random subspace, which generates i trees by random vector θ_i that obeys the independent identically distribution.

$$W_i\left(x,\theta_i\right), i=1,2,3,\cdots \tag{1}$$

All subtrees form the whole integrated tree model. For classification, the output of the RF model is the class voted by the most of the base classifiers. For regression, the output of the RF model is usually the average value predicted by all base classifiers. The algorithm flow of the random forest model is as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

(1) With the Bootstrap method, N sub-training sets $N_{ij}, i \in \{1, 2, 3, 4, \ldots, k\}$ are randomly selected from the training set with a total amount of W to form a CART tree for each training sub-sample.

(2) The regression random forest consists of i regression trees. The sub-nodes of each regression tree randomly select the number of splitting indexes n(n≤M) during splitting, where M is the number of indexes of the total sample. The optimal segmentation indexes are chosen according to the size of the measurement indexes for dividing.

(3) Repeat step (2) until all subtrees in the forest have been constructed.

(4) The final random forest is formed by i subtrees. The samples to be tested are introduced into the constructed random forest, and the final results are generated by averaging all the predicted valus. Its final decision function $P_{rf}(X)$ can be obtained from Equation (6):

$$P_{rf}(X) = \text{argmax}_Y \sum_{i=1}^{k} I\left(w\left(X, \theta_i\right) = Y\right) I(\cdot) Y_c \tag{2}$$

where, $w(X, \theta_i)$ is a single regression decision tree. $I(\cdot)$ is the index function to represent the total number of samples satisfying the formula; K is the number of subtrees to be built; Y is the target variable, which is interpreted as whether there is a default; θ_i is a random variable.

The decision result of the random forest depends on the training result of each subtree, and the selection of the splitting index determines the splitting standard. In regression random forest, minimum MSE is generally adopted, whose calculation is as follows:

$$min(A, s)\left[min(c_1)\sum_{x_i \in D_1(A, s)}(y_i - c_1)^2 + min(c_2)\sum_{x_i \in D_2(A, s)}(y_i - c_2)^2\right] \tag{3}$$

Among them, c_1 is the average output of data set D_1; c_2 is the average output of data set D_2. Each CART tree predicts the average value of the leaf nodes. The output of the RF is the average value of the predictions of all trees.

2.3 Extreme Gradient Boosting(XGBoost)

XGBoost algorithm is an extension of the Gradient Boosting Machine (GBM) algorithm, which is an optimization model with characteristics of the nonlinear model and tree model, and can complete regression and classification tasks at the same time. XGBoost algorithm is composed of multiple decision trees (CART), integrating decision trees to complete regression and classification tasks, and the cumulative regression values of all decision trees are the regression values of the model.

XGBoost adds a regular term to the original objective function of GBDT, thus speeding up the

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

convergence efficiency and reducing the risk of overfitting. The formula after transformation is as follows:

$$H^{(i)} = \sum_{i=1}^{n} \left(l\left(y_i, \widehat{y_i} \right) + \Omega\left(f_i \right) \right) \tag{4}$$

$$\Omega(f) = \gamma N + \frac{1}{2}\lambda \| O \|^2 \tag{5}$$

In the above formula, $l\left(y_i, \hat{y}_i \right)$ is the square error loss function between the predicted value $\hat{y}_i$ and y_i.

Equation (4) calculates the sum of complexity of all subtrees, where $Q(f)$ is the regularization term. N represents the number of leaf nodes in the subtree. A represents the punishment system value of the weight O of the leaf node; Y measures the difficulty of tree segmentation and is used to control tree growth.

The difference between XGBoost and GBDT lies in that the former is expanded by the Taylor formula's second derivative, thus accelerating the convergence rate of the function and improving the prediction accuracy of the model. The target function after transformation is:

$$H^{i} = -\frac{1}{2}\sum_{j=1}^{N} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma N \tag{6}$$

where $I_j \in \left\{ q\left(X_i \right) = j \right\}$, h_i is $h_i[l(\alpha)]$'', and g_i is $g_i[l(\alpha)]$'.

III. RESULTS OF EXPERIMENTS

1. Experiment Environment: Jupyter Notebook

TABLE 1: The main packages used in the experiments

Packages	Descriptions
Numpy	A package for scientific data operations
XGBoost	One of the ensemble learning algorithm packages
Sci-kit learn	Package of machine learning

2. Evaluation metrics: MAE, MSE, RMSE, and PTR

MAE represents the average absolute error between the groundtruths and the predicions. MAE directly calculates the average of residual error, and the calculation formula is as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{y_i}\right| \tag{7}$$

MSE represents the squared error between the groundtruths and the predicions. If the difference between the groundtruths and the predicions is greater than 1, the sampling error with a large difference between the groundtruths and the predicions will be further amplified. If less than 1, the error between the groundtruths and the predicions will be further reduced, and the calculation formula is shown as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2 \tag{8}$$

RMSE represents the sample standard deviation of residuals between predicted and observed values. Compared with MAE, RMSE penalized the sample corresponding to the predicted value that differed greatly from the true value.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2} \tag{9}$$

In addition to the above classical measures, we also introduce PT Ratio, which represents the Ratio of model performance to training time. Its calculation formula is defined as follows:

$$PTR = \frac{Perf\,ormance(\,score)}{TimeCost(\,sec)} \tag{10}$$

In the formula, Performance can be the measure of any evaluation model. Here, we choose the prediction precision of the model as the measure, and the larger the value is, the better the performance of the model is.

3. Validation Methods: Cross-Validation

To ensure the objectivity of model evaluation, Cross-Validation was adopted in this experiment. This method divides the sample set into M mutually exclusive sets and conducts m times of model training. After each model training, different subsets are used as test sets to obtain the current evaluation result. Finally, the average result of M times of test is taken as the return result and output. The cross-validation method avoids the potential error caused by a single model test to some extent and takes the average of multiple validation results as the final result, which can objectively reflect the model performance.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

Figure 1: Cross-Validation

3.1 Data Descriptions

This paper uses Javafoil's airfoil generator to generate NACA 4-bit airfoil data set. To represent shapes, each airfoil is discretized (normalized to unit chord length) at 101 cosine intervals to produce smooth upper and lower surfaces, with leading and trailing edges fixed at (0, 0) and (1, 0), respectively. Then, the different values of the upper surface points are yu_1-YU_15, and the different values of the lower surface points are YL_1-YL_15, which are used as airfoil parameters. For each airfoil, the lift (CL), drag (CD) and torque (Cm) coefficients are obtained under Reynolds number and Mach number conditions under different AoA conditions.

After checking, 2x15+6 column data were obtained after eliminating abnormal data. The first 15 columns consist of the y coordinates of the upper surface at a fixed X position, and the next 15 columns consist of the Y coordinates of the lower surface at the same X position. Then the three columns are respectively composed of AoA, Reynolds number and Mach number, and finally the output CL, CD and Cm values. This paper normalized all data by removing their mean values and scaling them to unit variance.

3.2 Comparisons of Different Algorithms

3.2.1 Tuning of XGBoost

As an ensemble learning algorithm with strong ensemble optimization ability, XGBoost can integrate multiple weak learners into one large strong learner. In the experimental process, considering that XGBoost has a higher improvement space compared with Random Forest, it means that XGBoost can obtain better model performance than Random Forest by adjusting parameters. Therefore, in this section, we will focus on the tuning process of XGBoost and see its performance on predicting one of the coefficients, CL.

First, we set the maximum depth of the tree as 6, the learning rate as 0.3, and the number of five-fold

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

cross-validation as 200 times for evaluation, visualizing the loss of training set and test set in these 200 times of cross-validation, as shown in Fig2. It can be found that the loss of training set and test set is consistent, indicating that XGBoost under this parameter has no fitting phenomenon and has strong generalization ability.

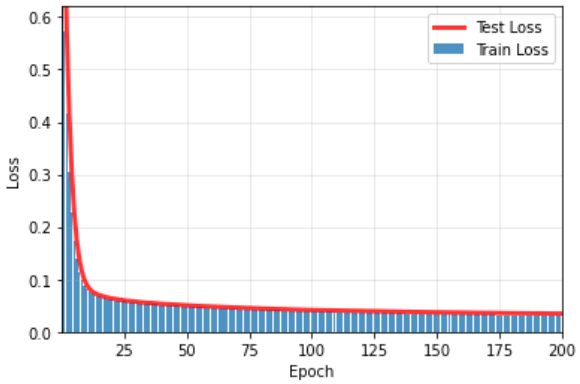


Figure 2: Train loss and test loss with increasing epoch

Therefore, we adjusted the prediction accuracy of the model under different learning rates, minimum loss function decline value, number of learners and tree maximum depth parameters several times, as shown in Fig3:

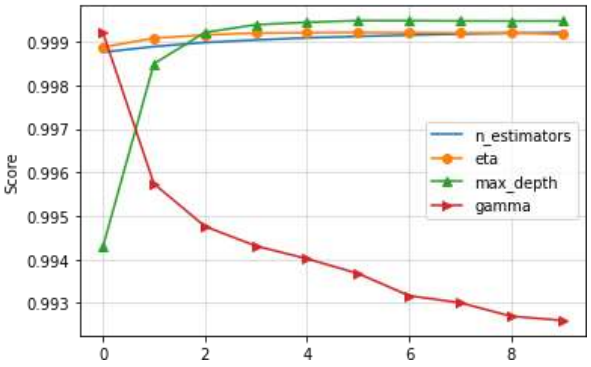


Figure 3: XGBoost performance under different parameters of different value

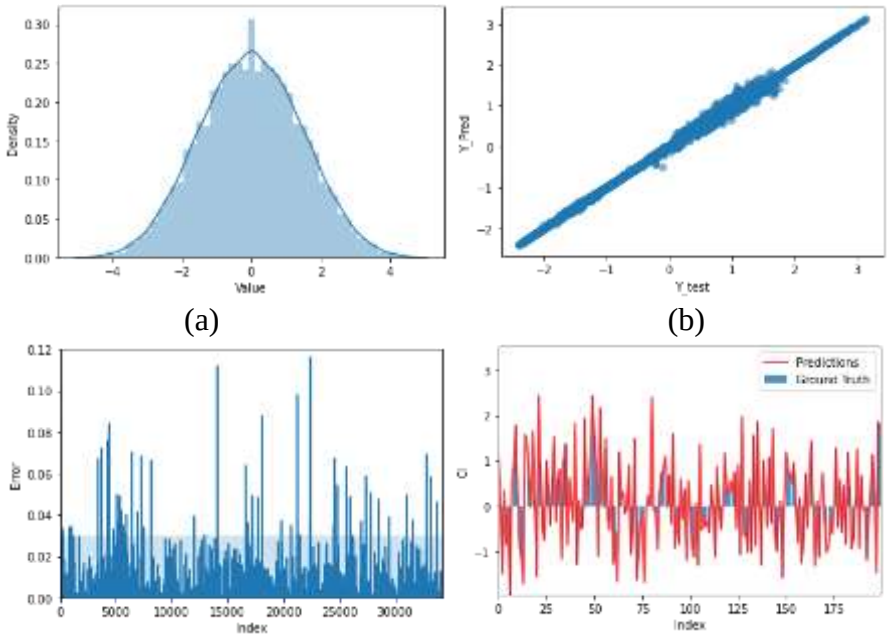
where x-axis represents the number of iterations. We have selected a total of 10 increasing parameter values for iteration. The numerical iteration lists of the four parameters are n_estimators =[200, 250, 300, 350, 400, 450, 500, 550, 600, 650], eta=[0.05, 0.1, 0.15, 0.2, 0.25, 0.3,0.35, 0.4, 0.45, 0.5], max_depth = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20], gamma = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]. After several pieces of training, we obtained the parameters of XGBoost with the best training performance on this data set, as is shown in Table 2. The optimal parameters of the following table can be summarized by observing the above Fig3.

TABLE 2: Optimal model training parameters

Learning Rate	Number of Estimators	Maximum Depth	Gamma
0.15	650	15	0

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

Finally, we trained the XGBoost model with the above group of parameters and compared some samples of predictions with groundtruths. The visualization results are shown in Fig4. It can be found that the predictions of the model trained with this group of parameters on different samples are consistent with the groundtruths, with high prediction accuracy (seeing Fig4(d), the groundtruths are bar graphs and the predictions are broken lines). The distribution of the predictions is roughly consistent with the distribution of the groundtruths.



(a) (b)

(c) (d)

Figure 4: Results of XGBoost after tuning

3.2.2 Comparisons

In this part, in order to demonstrate the superior performance of ensemble learning in regression more objectively, we compared the prediction accuracy between non-ensemble learning algorithms and that between ensemble algorithms, as shown in Table 3, 4.

According to the table, the MSE of support vector regression (SVR) is the lowest among non-integrated learning algorithms. Among the ensemble learning algorithms, XGBoost also achieves the desired optimality. At the same time, the training speed of XGBoost is faster than other ensemble learning algorithms. XGBoost adds a regular term in the cost function to control the complexity of the model. XGBoost takes a page from the random forest playbook and supports feature sampling, which not only prevents overfitting but also reduces computation, a feature that has been shown in experiments (with the highest PTR).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

The following errors are the average errors calculated by the cross-validation method. At the same time, the data set used in verification has no intersection with the training set, which can better reflect the strong generalization performance of ensemble learning.

TABLE 3: Performance and comparison of different non-ensemble learning algorithms

Regressor	MSE	Std	Regressor	MSE	Std
Linear Regression	0.169076	0.0192	Elastic Net	0.170728	0.0194
Ridge	0.170151	0.0190	SVR	0.073051	0.0134
Lasso	0.183093	0.0251	Bayesian Ridge	0.169061	0.0192

TABLE 4: Performance and comparison of different ensemble learning algorithms

Algorithms	Performance			
Random Forest	MAE	7.47528×10^{-3}	RMSE	2.11774×10^{-2}
	MSE	4.48480×10^{-4}	PTR	0.851
XGBoost	MAE	**6.81309×10^{-3}**	RMSE	**2.07202×10^{-2}**
	MSE	**4.29325×10^{-4}**	PTR	**9.255**
AdaBoost	MAE	0.22990	RMSE	0.28093
	MSE	0.07892	PTR	8.891
Gradient Boosting	MAE	0.06573	RMSE	0.09758
	MSE	9.52155×10^{-3}	PTR	4.818

IV. CONCLUSION AND FUTURE WORK

In this study, Random Forest(RF) and Extreme Gradient Boosting(XGBoost) are used to predict wing aerodynamic coefficients. The training data set is the NACA 4-bit airfoil data set generated by Javafoil. The airfoil is represented by 2x15 bit discrete airfoil surface coordinates, and the lift, drag and moment coefficients are predicted at different angles of attack, Reynolds number and Mach number. By comparing the training results of RF, XGBoost and non-ensemble learning algorithms, we find that the ensemble learning has higher prediction accuracy than the classical regression algorithm, and the prediction loss of RF and XGBoost is 1-2 orders of magnitude lower than that of other algorithms. Among them, XGBoost has the best performance, while RF has slightly lower prediction accuracy. Meanwhile, XGBoost has lower time complexity and higher generalization ability.

ACKNOWLEDGEMENTS

Funding Statement: This work is supported by the Innovation and Entrepreneurship Training Program for College Students from Hunan province(S202110537030), China.

REFERENCES

[1] SURESH S,OMKAR S N,MANI V, et al. Lift coefficient prediction at high angle of attack using recurrent neural network [J]. AerospaceScience and Technology, 2003, 7(8): 595-602.
[2] CARPENTER M,HARTFIELD R,BURKHALTER J.A comprehensive approach to cataloging missile aerodynamic performance using surrogate modeling techniques and statistical learning[C].//29th AIAA Applied Aerodynamics Conference,Honolulu,2011,27-30.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 1110-1120
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

[3] LIU X. Simulation of airfoil plunging aerodynamic parameter prediction based on neural network[J]. Computer Simulation. 2015,32(12):67-71.
[4] YUAN Z J,ZHANG G P,CUI M ,TANG W. Prediction of aerodynamic parameters based on neural network[J]. Aero Weaponry,2020,27(5):28-32.
[5] BALLA K,RUBEN S,OUBAY H, et al. An application of neural networks to the prediction of aerodynamic coefficients of aerofoils and wings[J]. Applied Mathematical Modelling, 2021, 96: 456-479.
[6] Y. Zhang, W. J. Sung, and D. N. Mavris, "Application of convolutional neural network to predict airfoil lift coefficient,"in AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2018, p. 1903.
[7] V. Sekar, Q. Jiang, C. Shu, and B. C. Khoo, "Fast flow field prediction over airfoils using deep learning approach, "Physics of Fluids, vol. 31, no. 5, p. 057103, 2019.]
[8] H. Chen, L. He, W. Qian, and S. Wang, "Multiple aerodynamic co-efficient prediction of airfoils using a convolutional neural network,"Symmetry, vol. 12, no. 4, p. 544, 2020.
[9] X. Hui, J. Bai, H. Wang, and Y. Zhang, "Fast pressure distributionprediction of airfoils using deep learning," Aerospace Science and Technology, vol. 105, p. 105949, 2020.
[10] X. Guo, W. Li, and F. lorio, "Convolutional neural networks for steady flow approximation, "in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 481 490.
[11] B. V. Dasarathy and B. V. Sheela, "A composite classifier system design: Concepts and methodology," in Proceedings of the IEEE, vol. 67, no. 5, pp. 708-713, May 1979, doi: 10.1109/PROC.1979.11321.
[12] Yoav Freund, Robert E Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, Volume 55, Issue 1, 1997, Pages 119-139, ISSN 0022-0000.
[13] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.
[14] Breiman L. Random forests. Ma chine Learning, 2001, 45(1):5 -32. [doi:10.1023/A:1010933404324].
[15] Rao Shanshan, Leng Xiaopeng.Random Forest credit evaluation based on combination of feature selection [J]. Computer system application, 2022, 31 (3): 345-350. The DOI: 10.15888 / j.carol carroll nki. Csa. 008371.
[16] Liu Jing-wen, Bai Jin-peng, WANG Jian, XU Ren, XU Wei-nan. Progress in Science and Technology of Water Resources and Hydropower,202,42(02):101-106.