A Review of Machine Learning Application in Aircraft Aerodynamics Development: Methods, Challenges and Prospects

Ziqi Li

Shenzhen College of International Education Shenzhen, China

s23007.li@stu.scie.com.cn

Abstract. Recently, with the development of AI technology and increasing requirement of Fluid Mechanics analysis, the utilization of machine learning (ML) in engineering is becoming an increasingly prominent research topic. In these engineering subjects, Fluid Mechanics is the one of the most discussed. Some researchers have already done diverse investigations in fusion of ML with Fluid Mechanics calculations. However, due to various limiting factors, the problems of relatively low accuracy, high price and large need of training data still exist in ML prediction results compared to Computational Fluid Dynamics (CFD) calculations. In this review, the present research in Fluid Mechanics and ML is presented. Achievements of ML application in aircraft optimization are briefly introduced with several examples. Then 3 well-developed ML aircraft aerodynamic performance analysis and optimization methods, which are neural networks (NNs), regression and clustering, are introduced. Their basic concepts and methodologies are analyzed with examples of applications. The benefits and drawbacks are compared, revealing that ML is tend to become another powerful tool of aerodynamics calculations in the near future. The review will provide the readers with a clearer understanding of ML development in aerodynamics.

Keywords: aerodynamics, machine learning, aircraft optimization, algorithms, review.

1. Introduction

Increasing globalization and cooperation nowadays has an urgent demand for higher quality in international and intercontinental transportation, which are mainly comprised of air transport. This places a challenge to current commercial transportation aircraft improvement, requiring better performance, higher efficiency and stronger sustainability, which makes aircraft optimization much more difficult. There must be innovations in different aspects of aircraft to improve their ability. One of the aspects that will essentially improve the overall performance of the aircraft is its aerodynamic design optimization, which is also a very hot topic in Fluid Mechanics research. For example, by innovating the design of the wings, important characteristics such as drag will be reduced. The development of design could be done by Computational Fluid Dynamics (CFD)-based optimization or surrogate-based optimization [1].

Plenty of achievements have been made by several researchers. Aelaei et al. [2] used CFD-based method and numerical simulations to optimize the configuration of clipped delta wings. They obtained the best design for highest efficiency and produced another optimization sample for air-launch-to-orbit aircraft. While in the field of morphing wing research, Cheng Gong and Bao-Feng Ma [3] have overcome the challenge of morphing aircraft shape optimization by using surrogate models and global sensitivity analysis, which is the first attempt ever reported. The research successfully gained configuration of the morphing wing under different speed interval that has the maximum lift-to-drag ratio and opened a door for further developments in morphing wings in aircraft. Viviani et al. [4] investigated the ability and performance of CFD designs for the latest and future supersonic aircraft. Comparing traditional methods with new RANS methods, providing an outlook of various future developments with high-speed aircraft in different condition, the research places a clear overview over the development of hypersonic aerodynamics under assistance of CFD analysis.

However, the time spent in traditional CFD calculations for optimization is relatively long, usually from hours to days [1]. This is often due to the large need for calculation, data collection and categorization in optimization process, which is strongly dependent on researcher's device used and their hardware sets such as CPU. This will not only cause problems of strict limitation for calculation environment, but also may elicit problems of high consumption of energy and low time-efficiency during optimization. Some researchers thus seek alternatives from currently fast developing Artificial Intelligence (AI) and Machine Learning (ML) technologies. By intaking data, ML can help researchers to derive corresponding predictions and information [5].

ML has had successful applications and proved helpful in fluid mechanics, its extension fields and other research majors. Afzal et al. [6] investigated high speed fluid flow from nozzles with assistance of ML. By utilizing K-means clustering, the corresponding input values were obtained when the output was clustered and matched with possible input value range. This helped researchers much easier to find which input would lead to possible output range. In addition, the application of Random Forest algorithm, which was trained and tested by experimental measurements, provided fitting estimations to experiment data. This reinforces the robustness of ML methods in fluid mechanics prediction. When it comes to a more specified topic, Bounds et al. [7] made a case study on ML prediction for solving SST k−ω Reynolds-averaged Navier–Stokes (RANS) turbulence model of a ground vehicle. It was found that Random Forest algorithm and XGB model are the best two methods to understand ground vehicle CFD analysis data and provide most accurate estimation in compare of 5 ML algorithms. This gives suggestion for ML methods choosing in future research of vehicle aerodynamic optimization. Apart from these, in other subjects such as medical, Yuvaraj and SriPreethaa [8] discussed the adhibition of a ML algorithm called Hadroop cluster in prediction of diabetes. They summarized this well-developed method and reinforced the importance of ML utilization in medical research.

Eventually, from former research it is found that the application of ML in aerodynamics is beneficial and promising, although the combination of these two subjects isn't as firm as in other fields. In the following content of this review, section 3 introduces some well-known ML algorithms with exact examples. Pros and cons of these algorithms are presented. In section 4, these pros and cons are brought into discussion of possible adhibition conditions, eliciting the vision of development of ML in aerodynamics. There will be a summary of the whole review in section 5.

2. Machine Learning Methodology

ML depends on the large amount of data stored in databases of aerodynamic principles and measurements. ML's ability to elicit information from databases is soon utilized and developed in creating aerodynamic models and predictions [9].

2.1. Neural Networks (NNs)

NN is designed to imitate the ability of the human brain using a computational modeling function. Although NNs may have diverse complexity and functions, such as Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN) with convolutional layers, Recurrent Neural Networks (RNN) with specialized units for long-term dependencies, their basic structures are similar. A basic structure diagram of an NN could be represented as in **Fig.1** [9].

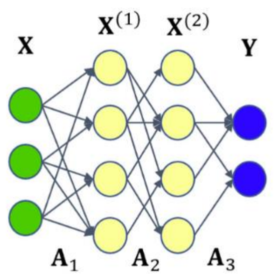

Fig. 1 A stucture diagram for data processing in a simplest neural network system

In **Fig.1**, $\mathbf{X}$ represents the input layer, which receives raw data as input of this network and is also map of the output layer $\mathbf{Y}$. The hidden layers, which are $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ in between input layers and output layers in **Fig.1**, are used to process and learn through complex patterns, determining the relationship between input and output. $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ are functions that have the detailed processes of processing nodes, represented in different colored circles, deal with the data they receive from the last layer. Each node receives input via connection with many other nodes in layers, represented by arrows in **Fig.1**. These connections have associated weights that determine their importance. An NNs may also include bias terms, which increase efficiency and flexibility.

NNs learn via supervised learning. Input will first pass through each layer and produce a set of predicted results. The results will be compared with actual results. Errors in these results will be quantified and fed back to adjusting weight and biases. Repetitions are needed until the system can produce relatively accurate predictions.

In aerodynamics, an NN is trained and tested mostly by CFD calculation results until the system can do relative accurate estimations of aerodynamics under the same condition. For example, Balla et al. [10] applied novel multi-output neural networks to a 2D and a 3D case of wing's aerodynamic coefficient prediction. It is shown that this specific NN has better accuracy compared to normal NNs in both predicting 2D airfoil and 3D wing. However, the accuracy is still hinged to CFD data. When the CFD data is more accurate, the prediction result is less accurate due to the increasing shocks in finer mesh analysis. Du et al. [11] combined surrogate modeling to a case study of an airfoil optimization. They have successfully generated a fast and efficient way of optimization using B-spline-based generative adversarial network model using specific parameters, proving the possibility of using NNs in aerodynamic as a rapid alternative to traditional methods such as CFD. The optimization result of NN involved method has only little difference to conventional CFD optimization result, provide another evidence for accessibility of using NNs in aircraft optimization.

In applications, NNs are fast, efficient and accurate if the training data is well selected. On the contrary, it still has limitations: the time and data needed for training may be a lot more than CFD calculations.

2.2. Regression

Regression was first introduced in statistics, with the function of predicting a continuous variable change by input value feature. For most of the cases, it involves functions and mapping. An algorithm will estimate a function according to scattered output that is the "best" fit line representing changing trend. **Fig.2** and **Fig.3**[9] show two diagrams explaining two kinds of regressions, linear regression and logistic regression.

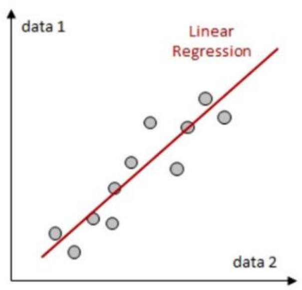

Fig. 2 A presentation of data trend summary in linear regression method

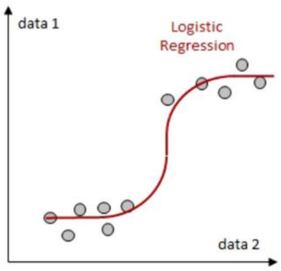

Fig. 3 A presentation of data trend summary in logistic regression method

From these diagrams it is clear to see that the sums of the bias of the points are kept the smallest to make sure that the line is representative and accurate. Apart from these two kinds of regression, there are many other kinds of regression: Polynomial regression, logistic regression and support vector regression etc. Yet their purposes are the same: To find the "best" fit line of continuous variables.

Regression can be combined with ML and learning from supervised learning. In the training stage, data is input so that the system can provide a regressed function for variables changing with other variables until the function is at its "best" fit condition. Then predicted data could be produced from the function with consideration of residual and standard error.

In aerodynamics, various regression methods can be combined and utilized for estimation. For instance, Weinmeister et al. [12] analyzed a regression-based polynomial chaos-kriging metamodel for quantifying uncertainties in aerodynamics designing. They take airfoils and engine nacelles as case studies, showing the high efficiency and accuracy of the model. Their investigation also illustrates that regression models have strict limitations on data's trend of changing, relatively low reliability and low adaptability. Compromises of accuracy and efficiency also must be considered. Teimourian et al. [13] compared other widely used regression algorithms, from complex Random Forest algorithm to simplest linear regression, on differences between aerodynamic performance estimation accuracy of an angled airfoil provided by these methods. It is found that under certain conditions the accuracy of Random Forest algorithm is surprisingly high, with highest R^2 value of 0.944 and MSE value of 0.0008. This proved that the regression methods are possible to have accurate predictions. However, the high accuracy depends on strict requirements of train/test data ratio for the method, which is still a limitation in regression method.

2.3. Clustering

Clustering can be classified into 5 types: Partitional Clustering, Hierarchical Clustering, Density-based Clustering, Grid-based Clustering and Model-based Clustering [14]. There are several differences between them. **Fig.4** [9] shows a diagram of K-Nearest neighbours, one of the clustering algorithms. In the diagram it is clearly shown that data within certain distance to a point are kept in a cluster

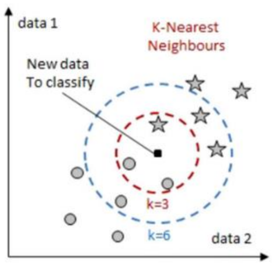

Fig. 4 Diagram of K-Nearest neighbours

2.3.1 Partitional Clustering

This is the clustering method that finds similarities between data. The center of the cluster is initialized first, then assignment of different data to different clusters is run based on the distances calculated between the data and the cluster-center. E.g. K-Means clustering, K-Modes clustering.

2.3.2 Hierarchical Clustering

This kind of clustering can have two approaches. In agglomerative approach, each data is considered as a cluster. Then the two nearest data are joined together to form a cluster. The process is repeated until a cluster with all data is formed. In divisive approach, everything is the opposite to agglomerative. A huge cluster including every data is separated layers by layers. E.g. ROCK, Echidna.

2.3.3 Density-based Clustering

In this kind of clustering, a random point of data is taken first. Then the area within a specific range is checked for its data density. Only the areas with high enough data density are recognized as clusters, while the other areas are used as partitions. In such a case the noisy data can be blocked from clusters. E.g. DBSCAN, DBCLASD

2.3.4 Grid-based Clustering

This type of clustering will create a grid by data from database. Then the density in each unit of the grid is calculated. If the density is above the limit value, the unit is used as a cluster. E.g. CLIQUE, STING.

2.3.5 Model-based Clustering

Model-based Clustering makes use of written mathematical models and make clusters from these set models. Different from other clustering, model-based clustering depends on combination of underlying probability and works out clusters from that. E.g. COBWEB.

In aerodynamics, although clustering is not as popular as conventional optimization skills, it performs its supportive functions in plenty of optimizations. For example, Zhao et al. [15] developed an optimization algorithm for vertical takeoff and vertical landing (VTVL) vehicles making use of K-means clustering algorithms, which is a partition clustering method. The sensitivity of data elements was categorized using the K-means clustering, which supported the gaining of linear and nonlinear constraint variables and design variables. Zhu et al. [16] made a case study on flying data of a commercial aircraft in purpose of discovering relationship between factors of fuel consumption, pilot behavior and flight factors. With assistance of improved K-means clustering method, it was much easier to investigate the relationships between these factors, providing data with lowest var of 0.125, which is quite reliable compared to original K-means method. They successfully developed an accessible optimized method of clustering in data analysis and proved the effectiveness of using clustering in data collection.

3. Discussions on Applications

In the texts above, three well-developed ML methods and their principles are introduced. More detailed features and applications are introduced in this section.

3.1. Discrepancies

NNs is the most distinct one among the three, it learns through supervised learning and provides numerical predictive data after learning and testing. It has high efficiency compared to other traditional calculation methods in aerodynamics such as CFD simulation. It only needs a few seconds to provide a set of predictive data while CFD needs hours of time. That’s why it is so attractive for researchers to develop. On the other hand, it has relatively lower accuracy and higher cost compared to simulation result. In Du et al.’s research [12], different models of NNs must be constructed to simulate and estimate parameter optimization under different mach number. They must divide the

Mach number ranges in [0.3, 0.6] and (0.6, 0.7] for relative accurate optimization. Also, there was still some differences between NN predicted data and CFD calculation result. Fortunately, recent research on NNs is focused on improving fidelity and they have already had some achievements in applications. It is believed that by combining NNs to other ML methods or improve the NNs algorithm to make it more relevant to the research topic are more possible to address the problem. Ko et al. [16] developed an application of neural networks in chemical problems, overcoming the limitations of considering electronic structures' global change, correctly illustrating accurate global charge distribution and broadens the opportunity for accessing high-accuracy machine learning methods.

NNs are closely dependent on data. It is found to be appropriate for situations when the data is plentiful, changing regularly and the problem is hard to understand in a specific model. When the data cannot fully cover most of the condition, or the disturbance in data is significant, NNs might not be the perfect choice for prediction.

Regression and clustering are two methods that are usually mixed up. Regression is the method that learns through supervised learning and produces a summative trend of data changing with parameters. Whereas clustering is an unsupervised learning ML method. It cannot be used to estimate any trend or result predictions but can only categorize sets of data. Regression is derived from statistical methods and can be combined with ML methods. Hence it is a very basic statistical skill and almost used in much research. For examples, Raul and Leifsson [17] applied Kriging regression model to estimate and restrict functions in aerodynamic airfoil optimization research and Ren et al. [18] utilized support vector regression and stochastic gradient descent together and developed a way for aero-engine optimization. While clustering is not that widely and directly used in aerodynamics optimizations due to its limitations of functionality in aerodynamics. It is more often used in problem division and data treatment.

Regression has strict requirements on data extensiveness. When using a regression for aerodynamic optimization, the data for the regression model must cover the parameters well. In addition, the model needs to be generalized well, with a good balance between complexity and generalization capability so that the regression will not overfit.

On the other hand, clustering in aerodynamics optimization is a useful tool. Although it cannot be directly used in parameter prediction, it is very helpful when there's need of simplifying large amounts of complex data, revealing trend between parameters or dividing a composite problem into sub-problems.

3.2. Envisages

From former examples, the strong capabilities of ML methods in various fields are fully represented. However, the application of ML in aerodynamics is still developing. In the combination of ML with traditional CFD calculations and aerodynamics optimization models, there still exists flaws such as relatively low fidelity, huge time and financial costs to construct and train ML models and low utilization efficiency etc.

For low accuracy problem, there are plenty of opportunities to explore newer models in aerodynamics to solve the problem. For lowering time and financial cost, it is hoped that as the development and popularization of AI technology, it could be easier to utilize these powerful ML techniques in research. For increasing utilization efficiency, researchers may combine different ML methods to more complex problems and optimize the algorithm of ML models. Using the multi-system model, the result will be obtained with much higher efficiency compared to former research. Gradually, this powerful tool will be better used in aerodynamics.

It is believed that with the help of ML, the important aerodynamic design and optimization work will benefit from its high efficiency and incredible adaptability. With these essential improvements, the productivity of optimization work will be significantly increased, which can push the advancement of aircraft performance and eventually in air transport worldwide.

4. Conclusion

The growth of ML applications and aerodynamics itself have had a huge impact on the research trying to combine these two subjects. A lot of researchers are attracted to the high efficiency and productivity of ML methods compared to traditional optimization models. The goal of research on the topic is also investigating how to utilize ML's advantages and serve it as a better alternative to conventional methods such as CFD. Among these popular ML techniques, the review explored three of them: Neural Networks, Regression and Clustering.

The review comprehensively summarizes the three measures. These methods have different working principles and play different roles in aerodynamics optimization works. NNs depends on a logic network, works out predictive data via learning process. Regression utilizes statistical formulas, summarize the trend of data changing with parameters. Clustering observes the differences between data and classifies them by learning. Various research examples are given and analyzed the role of ML in each of them.

When it comes to application of ML in aerodynamics, it is noticeable that ML will not be a replacement of CFD. Instead, it should be a powerful tool, an alternative algorithm maintaining the robustness and high efficiency in CFD calculations. In an optimization case, a well-trained ML algorithm can quickly make predictions of data and minimize the range of possible optimal data instead of depending on long CFD calculations. It can also help with data collection and classification, which minimizes the possibility of mistakes in data trend summarization and optimization failure.

Continuing with the introduction of three ML methods' principles, their state-of-the-art application situations, suitable utilization conditions and envision in their future developments are elicited. Although the limitations and drawbacks exist, ML will be a more powerful tool in aerodynamics designing and optimizing as the AI technology popularizes.

References

[1] Li, J., Zhang, M., Data-based approach for wing shape design optimization, Aerospace Science and Technology, 112 (2021), 106639J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Aelaei, M., Karimian, S., Ommi, F., Sensitivity Analysis and Optimization of Delta Wing Design Parameters using CFD-Based Response Surface Method, Journal of Applied Fluid Mechanics, 12 (6) (2019), 1885-1903

[3] Gong, C., Ma, B., Shape Optimization and Sensitivity Analysis of a Morphing-Wing Aircraft., Int. J. Aeronaut. Space Sci., 20 (2019), 57–69

[4] Viviani, A., Aprovitola, A., Pezzella, G., Rainone, C., CFD design capabilities for next generation high-speed aircraft, Acta Astronautica, 178 (2021), 143-158

[5] Mahesh, B., Machine Learning Algorithms - A review, International Journal of Science and Research, 9 (1) (2020), 381-386

[6] Afzal, A., Aabid, A., Khan, A., Khan, S. A., Rajak, U., Verma, T. N., Kumar, R., Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows, Aerospace Science and Technology, 107 (2020), 106318

[7] Bounds, C.P., Desai, S., Uddin, M, Enhancing CFD Predictions with Explainable Machine Learning for Aerodynamic Characteristics of Idealized Ground Vehicles. Vehicles 6 (2024), 1318-1344

[8] Yuvaraj, N., SriPreethaa, K.R. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Comput 22 (Suppl 1) (2019), 1–9

[9] Le Clainche, S., Ferrer, S., Gibson, S., Cross, E., Parente, A., Vinuesa, R., Improving aircraft performance using machine learning: A review, Aerospace Science and Technology, 138 (2023), 108354

[10] Balla, K., Sevilla, R., Hassan, O., Morgan K., An application of neural networks to the prediction of aerodynamic coefficients of aerofoils and wings, Applied Mathematical Modelling, 96 (2021), 456-479

[11] Du, X., He, P., Martins, J.R.R.A., Rapid airfoil design optimization via neural networks-based parameterization and surrogate modeling, Aerospace Science and Technology, 113 (2021), 106701

[12] Weinmeister, J., Gao, X., Roy, S., Analysis of a Polynomial Chaos-Kriging Metamodel for Uncertainty Quantification in Aerodynamics, AIAA Journal, 57 (6) (2019), 2280-2296

[13] Teimourian, A., Rohacs, D., Dimililer, K., Teimourian, H., Yildiz, M., Kale, U., Heliyon, 10 (8) (2024),

[14] Ghosal, A., Nandy, A., Das, A., Goswami, S., Panday, M., A Short Review on Different Clustering Techniques and Their Applications, Emerging Technology in Modelling and Graphics (2020), 69-83

[15] Zhao, J., He, X., Li, H., Lu, L., An adaptive optimization algorithm based on clustering analysis for return multi-flight-phase of VTVL reusable launch vehicle, Acta Astronautica, 183 (2021), 112-125

[16] Zhu,Q., Pei, J., Liu, X., Zhou, Z., Analyzing commercial aircraft fuel consumption during descent: A case study using an improved K-means clustering algorithm, Journal of Cleaner Production, 223 (2019), 869-882

[17] Ko, T.W., Finkler, J.A., Goedecker, S. et al. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer, Nat Commun 12 (1) (2021), 39

[18] Raul, V., Leifsson, L., Surrogate-based aerodynamic shape optimization for delaying airfoil dynamic stall using Kriging regression and infill criteria, Aerospace Science and Technology, 111 (2021), 106555

[19] Ren, L.H., Ye, Z.F., Zhao, Y.P., A modeling method for aero-engine by combining stochastic gradient descent with support vector regression, Aerospace Science and Technology, 99 (2020), 10577