

ПРИМЕНЕНИЕ АНСАМБЛЕВЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ СУРРОГАТНЫХ МОДЕЛЕЙ В ЗАДАЧАХ ПРЕДВАРИТЕЛЬНОГО ПРОЕКТИРОВАНИЯ ПРОФИЛЯ КРЫЛА ЛЕТАТЕЛЬНОГО АППАРАТА¹

Петров Д.А., Голев А.В., Московцев А.М.

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия
daniel.petrov@ipu.ru, oiw23@mail.ru, moskovcev@ipu.ru

Аннотация. В работе исследованы методы построения суррогатных моделей с использованием ансамблевых методов машинного обучения. Суррогатные модели позволяют значительно сократить время, затрачиваемое на поиск рациональных геометрических параметров крыла летательного аппарата. В работе рассмотрена задача поиска рациональных аэродинамических характеристик для профиля крыла летательного аппарата. Обучающий массив данных был сгенерирован комплексом XFOIL - программным обеспечением для анализа характеристик аэродинамических профилей на дозвуковых скоростях. В результате расчетных экспериментов выяснилось, что наилучшей регрессионной моделью-ансамблем для данной задачи является Extra Trees Regressor – ансамбль, основанный на построении нескольких моделей деревьев решений.

Ключевые слова: машинное обучение, летательный аппарат, регрессия, ансамбль моделей, суррогатная модель, аэродинамический профиль.

Введение

Современные технологии и методы машинного обучения играют ключевую роль при проектировании сложных инженерных систем, таких как летательные аппараты. Одна из задач аэродинамики заключается в подборе оптимальной конструкции аэродинамического профиля крыла летательного аппарата. Данная задача требует учета множества факторов таких как устойчивость, управляемость летательного аппарата, энергоэффективность и другие. Основными характеристиками аэродинамического профиля являются коэффициент сопротивления и коэффициент подъемной силы крыла. Оптимальная комбинация данных характеристик позволяет повысить эффективность и безопасность полета летательного аппарата. В работах [16, 17] показано, что на первых этапах предварительного проектирования решать данную задачу прямыми методами вычислительной аэродинамики очень затратно с точки зрения вычислительно-временного ресурса. Один из способов решения данной проблемы заключается в построении суррогатной модели на основе точной аэродинамической модели. Суррогатные модели, также известные в литературе как метамодел, представляют собой приближенные модели, которые воспроизводят поведение сложных систем со значительно меньшими вычислительными затратами. Первым этапом построения суррогатной модели генерируется обучающий массив данных, что может быть реализовано с использованием численных методов, таких как вычислительная гидродинамика (CFD), которая требует значительных вычислительных ресурсов и времени для решения уравнений Навье-Стокса. Для создания устойчивой суррогатной модели необходимо собрать данные о поведении точной модели при различных условиях эксплуатации, что включает проведение множества симуляций с вариацией геометрических параметров профиля крыла и условий обтекания. С использованием собранных данных строится суррогатная модель, которая может быть реализована применением различных методов машинного обучения, таких как линейная и нелинейные модели регрессии, гауссовские процессы, нейронные сети, а также регрессионные модели-ансамбли. После стадии обучения суррогатная модель проходит этап валидации и тестирования для оценки её точности и надежности. Это включает сравнение предсказаний суррогатной модели с результатами точной модели на новых данных, которые не использовались при обучении. Использование суррогатной модели предоставляет значительные преимущества, так как данные модели значительно быстрее выполняют расчеты по сравнению с точными моделями, что позволяет ускорить процесс проектирования основных геометрических параметров аэродинамического профиля, при этом суррогатные модели требуют значительно меньше вычислительных ресурсов, что позволяет эффективно использовать доступные вычислительные мощности. При этом, суррогатные модели легко адаптируются к изменениям в проектных параметрах и могут быть использованы для оперативного исследования множества допустимых конструктивных решений.

¹ Исследование частично выполнено за счет гранта Российского научного фонда (проект № 24-19-00430)

1. Построение суррогатной модели аэродинамического профиля

Аэродинамические профили имеют разнообразные формы, включая симметричные и несимметричные, толстые и тонкие. Эти профили могут быть оптимизированы для определенных условий полета и различных типов летательных аппаратов. На рисунке 1 изображена схема базовых геометрических характеристик для аэродинамического профиля крыла летательного аппарата. К основным параметрам профиля относятся: Относительная вогнутость профиля (Max_c), координата максимальной вогнутости профиля (Max_c_p) и относительная толщина профиля (Th).

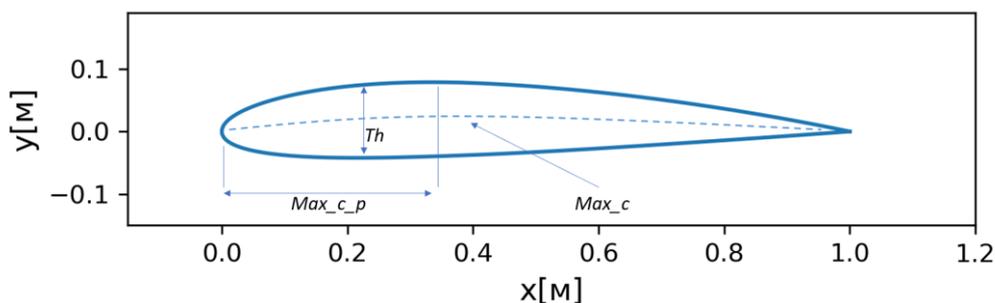


Рис. 1. Схема аэродинамического профиля крыла

Существует множество методов для построения геометрии аэродинамического профиля [20]. Некоторые из них основываются на явных уравнениях кривых, задающих верхние и нижние контуры профиля, тогда как другие используют параметрически заданные кривые [21]. В данной работе для формализации геометрии аэродинамического профиля были выбраны четырехзначные профили NASA [22].

2. Формирование обучающего массива

Для построения обучающего массива данных для суррогатной модели использовался программный комплекс XFOIL [23]. Входные параметры модели:

1. Число Маха - величина, используемая для определения скорости объекта относительно скорости звука в рассматриваемой среде
2. Относительная вогнутость профиля – отношение максимального значения вогнутости профиля к хорде.
3. Положение максимальной вогнутости профиля – координата на хорде, где профиль обладает максимальной вогнутостью.
4. Относительная толщина профиля – отношение максимальной толщины профиля к хорде.
5. Угол атаки - угол между направлением вектора скорости набегающего потока и характерным продольным направлением, выбранным на теле.

Выходные параметры:

1. Коэффициент лобового сопротивления (C_x) - величина, используемая для описания силы сопротивления, действующей на объект, движущийся в потоке жидкости или газа
2. Коэффициент подъемной силы (C_y) - безразмерная величина, используемая для описания подъемной силы, генерируемой аэродинамическим профилем.

В таблице 1 приведены значения входных параметров задачи для построения обучающего массива данных

Таблица 1. Входные параметры задачи для формирования массива данных для обучения модели

Число Маха (M)	[0.1, 0.225, 0.35, 0.475, 0.6]
Относительная вогнутость профиля (Max_c)	[10%, 50%, 90%]
Положение максимальной вогнутости профиля (Max_c_p)	[20%, 50%, 80%]
Относительная толщина профиля (Th)	[10%, 14%, 18%]
Угол атаки (α)	[0, 3, 6, 9, 12]

В результате расчетов был построен обучающий массив из 675 расчетных точек с двумя выходными параметрами – коэффициентом сопротивления и подъемной силы. На основе данного массива были построены две суррогатные модели, основанные на ансамбле Extra Trees Regressor.

3. Построение суррогатной модели с использованием моделей-ансамблей

Ансамбли моделей представляют собой метод машинного обучения, который объединяет несколько отдельных моделей для повышения их точности и общей предсказательной способности [1]. Такой подход основан на идее, что сочетание нескольких моделей может компенсировать ошибки отдельных моделей и улучшить общие результаты. Ансамблирование является одним из ключевых методов в машинном обучении и находит широкое применение в различных задачах таких как классификация, регрессия и кластеризация. К основным типам ансамблей относятся:

- Bagging (Bootstrap Aggregating) [3]: Этот метод заключается в создании нескольких версий модели путем обучения их на разных выборках данных, полученных методом бутстрепа (случайной выборки с возвращением). Затем предсказания этих моделей объединяются, обычно путем усреднения (для регрессии) или голосования (для классификации). Примером является метод случайных лесов (Random Forest), который объединяет множество деревьев решений.
- Boosting [3,4]: В этом методе модели обучаются последовательно, каждая следующая модель пытается исправить ошибки предыдущих. Самый известный алгоритм boosting — AdaBoost, который адаптивно настраивает веса для каждого примера в обучающем наборе данных. Другие примеры алгоритмов - Gradient Boosting и XGBoost.
- Stacking [4]: Этот метод объединяет несколько базовых моделей путем обучения метамодели на их предсказаниях. Базовые модели делают предсказания, которые затем используются в качестве входных данных для метамодели. Это позволяет выявить наиболее значимые предсказания и улучшить общую точность.
- Voting [4]: В этом методе несколько моделей обучаются на одном и том же наборе данных, и их предсказания комбинируются путем голосования (для классификации) или усреднения (для регрессии). Существует два типа голосования: жесткое (majority voting) и мягкое (с учетом вероятностей предсказаний).

Ансамблирование моделей является эффективным подходом для задач регрессии, особенно когда имеется большой объем данных. Большое количество данных позволяет лучше обучить базовые модели и обеспечить более надежные предсказания. Объединение предсказаний нескольких моделей помогает уменьшить ошибки и повысить точность [6]. Ансамблирование снижает риск переобучения, так как компенсирует ошибки отдельных моделей. Ансамбли моделей могут использоваться для различных типов задач и с разными алгоритмами машинного обучения. В данной работе исследовались следующие модели ансамблей-регрессоров:

- AdaBoost Regression, или адаптивная бустинговая регрессия, представляет собой метод машинного обучения, который улучшает производительность регрессионных моделей за счет последовательного обучения относительно слабых моделей регрессии. Алгоритм адаптивно настраивает процесс обучения таким образом, чтобы каждая последующая модель исправляла ошибки, допущенные предыдущими моделями. [9]
- Bagging Regressor - представляет собой метод машинного обучения, основанный на технике бэггинга (Bootstrap Aggregating), применяемой к задачам регрессии. Он использует ансамбль базовых регрессионных моделей, чтобы улучшить общее качество предсказания [3].
- Extra Trees Regressor, или регрессор на основе экстремально случайных деревьев, является одним из методов ансамблевого обучения, применяемым в задачах регрессии. Этот метод аналогичен алгоритму Random Forest, но с одним отличием: для построения деревьев решений используются случайные выборки данных [12,13]
- Градиентный бустинг (Gradient Boosting) — метод машинного обучения, который используется в задачах регрессии для построения ансамбля моделей, таких как деревья решений [15].
- Гистограммный градиентный бустинг (Histogram Gradient Boosting) - оптимизированная реализацию градиентного бустинга, позволяет значительно ускорить обучение модели без ущерба качества предсказания.

В таблице 2 приведены результаты вычислительных экспериментов со значениями коэффициента детерминации (R^2) и величиной среднеквадратичной ошибки ($RMSE$) для каждой модели. В результате экспериментов выяснилось, что наилучшей моделью для данной задачи является модель Extra Trees Regressor. Вектор входных параметров состоял из числа Маха (M), относительной вогнутости профиля

(Max_c), положения максимальной вогнутости профиля (Max_c_p), относительной толщины профиля (Th) и угла атаки (α). В качестве целевых значений использовались коэффициент подъемной силы (CL) и коэффициент сопротивления (CD), регрессионные модели строились отдельно для каждого целевого значения:

$$\begin{cases} [M, Max_c, Max_c_p, Th, \alpha] \rightarrow CL \\ [M, Max_c, Max_c_p, Th, \alpha] \rightarrow CD \end{cases} \quad (1)$$

Таблица 2. Результаты вычислительных экспериментов с моделями регрессоров

Модель регрессора	CL		CD	
	R^2	RMSE	R^2	RMSE
Extra Trees Regressor	0.999285	0.01214	0.996647	0.000458
Random Forest Regressor	0.999173	0.013051	0.995925	0.000505
Bagging Regressor	0.998974	0.01454	0.994932	0.000563
HistGradient Boosting Regressor	0.997375	0.023259	0.979783	0.001125
Gradient Boosting Regressor	0.992073	0.040418	0.94656	0.001828
AdaBoost Regressor	0.941046	0.110221	0.582137	0.005112

4. Верификация регрессионной модели

На рисунке 2 представлены графики зависимости величины коэффициента подъемной силы CL при разных значениях величины угла атаки (α) для профиля NACA3312 для величины числа Маха равной 0.2. Каждая из моделей регрессии сравнивается с точным значением величины коэффициента подъемной силы CL , вычисленной с использованием программы XFOIL. С использованием ансамбля ExtraTreesRegressor на выходе модели образуется наиболее гладкое распределение величины CL , которое наилучшим образом согласуется с точным решением.

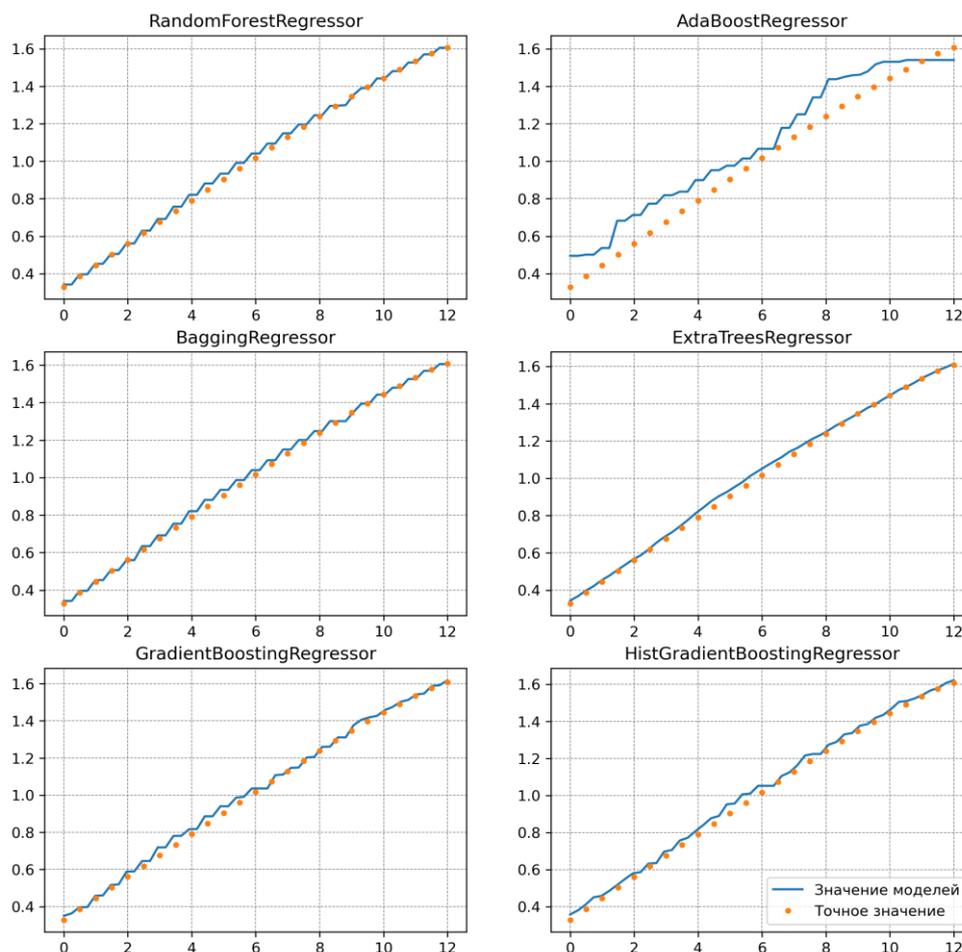


Рис. 2. Сравнение точного значения CL и значения CL для каждой модели регрессоров

5. Пример использования суррогатной модели для решения задач аэродинамики

В качестве примера, решим следующую задачу: необходимо подобрать аэродинамический профиль, обладающий следующими характеристиками и ограничениями:

$$\begin{cases} CL = 0.7 \\ CD \rightarrow \min \\ M = 0.2 \end{cases} \quad (2)$$

Значение CL для профиля должно быть равно 0.7, число Маха равно 0.2 при минимально возможном коэффициенте сопротивления CD . На рисунке 3 приведены графики основных геометрических характеристик профиля – относительной вогнутости (Max_c) и положения относительной вогнутости (Max_c_p) при постоянной относительной толщине профиля (Th) равной 12%. Визуальный анализ графиков показывает, что для каждого значения числа Маха (M) существует значение положения относительной вогнутости с минимальной величиной коэффициента сопротивления CD .

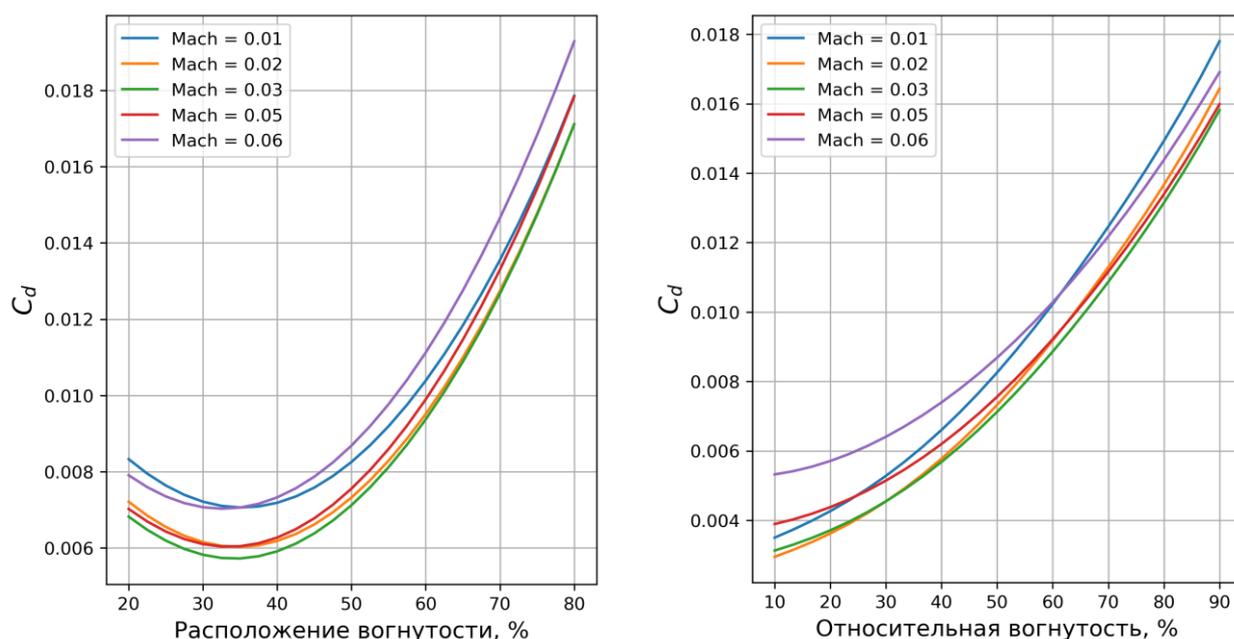


Рис. 3. Относительная вогнутость профиля и её относительное расположение на профиле

Для определения точного значения характеристик профиля воспользуемся известным оптимизационным алгоритмом NSGA- II (Non-dominated Sorting Genetic Algorithm II) [18, 19]. В результате поиска оптимальной конфигурации профиля на выходе задачи имеем три профиля-кандидата – см. таблицу 3. Также в таблице приведены верификационные значения CL_{ver} , CD_{ver} , полученные с помощью точных методов, ошибка в вычислениях составляет в среднем один процент. На рисунке 4 представлены геометрии профилей, удовлетворяющие условию (2). Каждый из полученных профилей с высокой степенью точности удовлетворяет условию поставленной задачи, при этом в зависимости от конструктивных ограничений у проектировщика появляется возможность выбрать наиболее подходящую геометрию. Более тонкий профиль NASA 3710 имеет более высокое аэродинамическое качество равное 76.9, в то время как качество более надежного профиля с точки зрения прочности NASA 1417 составляет 65.9.

Таблица 3. Результаты вычислительных экспериментов с моделями регрессоров

camber	camber_position	thickness	alpha	cl	cd	cl_ver	cd_ver	cl_dif%	cd_dif%
3	7	10	1.82	0.6805	0.00885	0.682114	0.008752	0.24%	1.12%
2	3	14	3.923	0.6713	0.00964	0.677442	0.009716	0.91%	0.78%
1	4	17	4.35	0.6729	0.01021	0.67446	0.01027	0.23%	0.59%

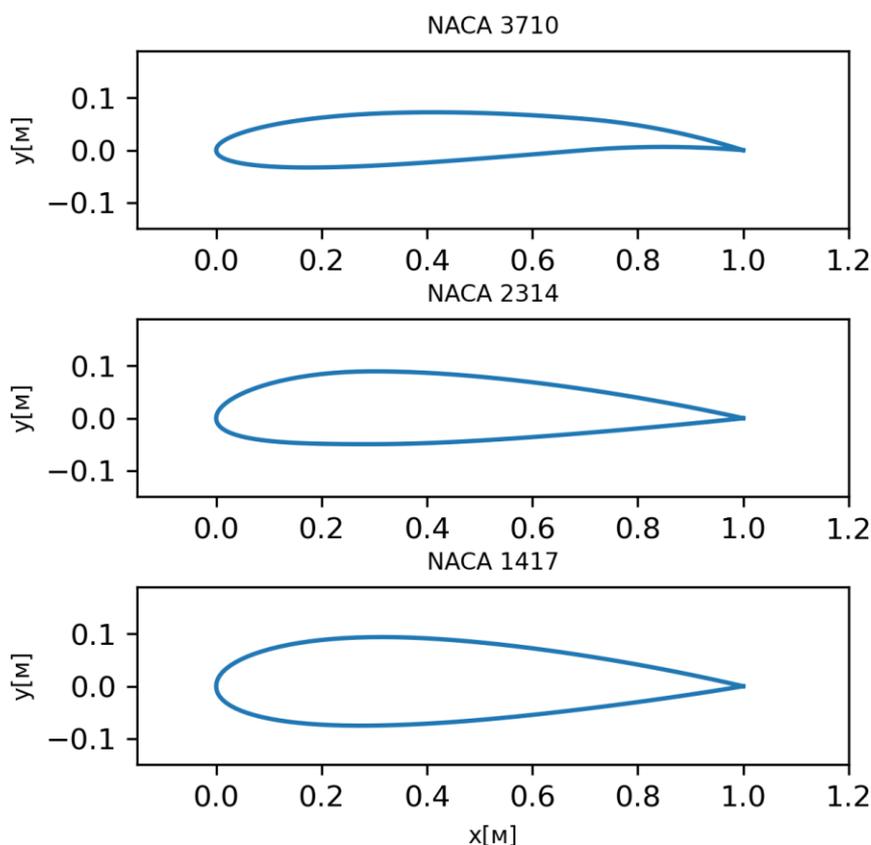


Рис. 4. Геометрия профилей, удовлетворяющих условию поставленной задачи

6. Заключение

Основываясь на обучающем множестве, сформированном с применением численных методов в аэродинамике, построена и верифицирована суррогатная модель для расчета основных характеристик аэродинамического профиля на основе регрессионной модели-ансамбля Extra Trees Regressor. В результате исследования были выявлены нелинейные взаимосвязи между геометрическими параметрами профиля и его основными аэродинамическими характеристиками, такими как подъемная сила и аэродинамическое сопротивление. Верификационные исследования показали, что модель в высокой степени точности согласуется с точными значениями основных аэродинамических характеристик полученными методами вычислительной аэродинамики.

Литература

1. *Kessy Salvatory, Sherris Michael, Villegas Andrés and Ziveyi Jonathan.* Mortality Forecasting Using Stacked Regression Ensembles (April 10, 2021). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3823511>.
2. *Devi M. Shyamala, Swathi P., Upadhyay Shubham Santosh, Sah Nitesh Kumar, Budhia Ankita, Chowdary Vamsi, Srivastava Sagar and Rohella Manshi.* Feature Predominance Ensemble Inquisition towards Liver Disease Prediction using Machine Learning (May 9, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3842561>.
3. *Colucci Stefano and Brandolini Dario.* Backtesting Value-at-Risk: A Comparison between Filtered Bootstrap and Historical Simulation (November 28, 2011). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.1965377>.
4. Computational Statistics & Data Analysis Volume 38 Issue 428 February 2002 pp 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
5. *Hvala Nadja, Mlakar Primož, Grašič Boštjan, Božnar Marija Zlata, Perne Matija and Kocijan Juš.* Surrogate Tree Ensemble Model Representing 2D Population Doses Over Complex Terrain in the Event of a Radiological Release Into the Air. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4181242>.
6. *Grushka-Cockayne Yael, Jose Victor Richmond and Lichtendahl Kenneth C.* Ensembles of Overfit and Overconfident Forecasts (August 18, 2015). Darden Business School Working Paper No. 2474438, Available at <http://dx.doi.org/10.2139/ssrn.2474438>.
7. *Sharafati A., Haji Seyed Asadollah S.B., Motta D. & Yaseen Z.M.* (2020). Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. Hydrological Sciences Journal, 65(12), 2022–2042. <https://doi.org/10.1080/02626667.2020.1786571>.

8. *Ita Kevin and Prinze Joyce*. Machine Learning for Skin Permeability Prediction: Random Forest and XGBoost Regression. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4415496>.
9. *Patil Sangram and Patil Aum and Phalle Vikas M*. Life Prediction of Bearing by Using Adaboost Regressor (December 13, 2018). Proceedings of TRIBOINDIA-2018 An International Conference on Tribology, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3398399>.
10. *Collins M., Schapire R.E. & Singer Y*. Logistic Regression, AdaBoost and Bregman Distances. *Machine Learning* 48, 253–285 (2002). <https://doi.org/10.1023/A:1013912006537>.
11. *Breiman L*. Using Iterated Bagging to Debias Regressions. *Machine Learning* 45, 261–277 (2001). <https://doi.org/10.1023/A:1017934522171>.
12. *Solomatine D.P., & Dulal K.N.* (2003). Model trees as an alternative to neural networks in rainfall—runoff modelling. *Hydrological Sciences Journal*, 48(3), 399–411. <https://doi.org/10.1623/hysj.48.3.399.45291>.
13. *Martiello Mastelini S, Nakano FK, Vens C, de Leon Ferreira de Carvalho ACP*. Online Extra Trees Regressor. *IEEE Trans Neural Netw Learn Syst*. 2023 Oct;34(10):6755-6767. doi: 10.1109/TNNLS.2022.3212859. Epub 2023 Oct 5. PMID: 36269923.
14. *McCarty Dakota and Kim Hyun Woo*, Exploring Risky Behaviors and Demographics: A Hybrid Extra Trees Regressor Approach for Road Accident Frequency Analysis. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4416105>.
15. *Richard S. Zemel and Toniann Pitassi*. 2000. A gradient-based boosting algorithm for regression problems. In Proceedings of the 13th International Conference on Neural Information Processing Systems (NIPS'00). MIT Press, Cambridge, MA, USA, 675–681.
16. *Московцев А.М., Петров Д.А.* Применение суррогатных моделей для предварительного проектирования несущих поверхностей летательного аппарата // ИТ-Стандарт. 2024, №1, С. 4-12.
17. *Пейгин С.В., Пуцун Н.А., Болсуновский А.Л., Тимченко С.В.* Оптимальное аэродинамическое проектирование крыла широкофюзеляжного дальнемагистрального самолета, *Вестн. Томск. гос. ун-та. Матем. и мех.*, 2018, № 51, С. 117–129.
18. *Deb K., Pratap A., Agarwal S., and Meyarivan T*. A fast and elitist multiobjective genetic algorithm: nsga II. *Trans. Evol. Comp*, 6(2):182–197, April 2002. URL: <http://dx.doi.org/10.1109/4235.996017>, doi:10.1109/4235.996017.
19. *Verma S., Pant M. & Snasel V*. Web service location-allocation using discrete NSGA-II with matrix based genetic operations and a repair mechanism. *J Ambient Intell Human Comput* 14, 14163–14187 (2023). <https://doi.org/10.1007/s12652-023-04625-6>.
20. *lv Pengjie & Pan Shenghu*. Research on NURBS Curve Interpolation Algorithm Based on Mline-Gear. *International Journal of Computer Science and Information Technology*, 2024, pp 84-94.
21. *Khurana Manujit, Winarto Hadi & Sinha Arvind*. Airfoil Geometry Parameterization Through Shape Optimizer and Computational Fluid Dynamics, 2008.
22. NACA2412 airfoil based method for design and aerodynamic analysis of small HAWT using modified BEM approach Devashish Jha, Saket Saurabh Sci. Tech. Energ. Transition, 2023.
23. *Drela Mark*, XFOIL: An Analysis and Design System for Low Reynolds Number Airfoils, Conference on Low Reynolds Number Airfoil Aerodynamics, University of Notre Dame, 1989.