International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 21s, 2025
https://theaspd.com/index.php

Prediction Of Air Quality Index Affected By Vehicle Emissions Using Machine Learning

Maulik Dhamecha[1], Urvi Dhamecha[2], Hansa Vaghela[3], Vijay Vyas[4], Hema Dhadhal[5], Bhaveshkumar Kathiriya[6]
[1]Department of Computer Engineering, V.V.P. Engineering College, Rajkot, Gujarat, India.
[2]Department of Computer Engineering, Marwadi University, Rajkot, Gujarat, India.
[3]Department of Computer Engineering, Marwadi University, Rajkot, Gujarat, India.
[4]Department of IT, V.V.P. Engineering College, Rajkot, Gujarat, India.
[5]Department of IT, L.E. College, Morbi, Gujarat, India.
[6]Department of EC, AVPTI, Rajkot, Gujarat, India.

Abstract
As a reaction to the increasing population everywhere globally and the rapid process of urbanization, development and manufacturing industries have increased. Growth has, however, increased levels of pollution, thus contributing to vast environmental issues like global warming. Out of all the sources of pollution, car emissions have a crucial role in affecting the condition of the atmosphere, thus making air pollution a primary cause of climate change.
This research tries to examine the effect of vehicle emissions on air pollution in India and forecast the Air Quality Index (AQI) using machine learning methods. This research is interested in learning about the effect of vehicle pollutants on the increase and decrease in AQI. Machine learning algorithms are used to predict AQI from vehicle emissions. Different prediction models are used to compare their efficacy in determining air quality. The research mostly utilizes information regarding vehicle pollution in India and its effect on AQI levels.
The research proves the capability of machine learning to forecast AQI based on vehicle emissions. Varying algorithms yield varying outcomes when it comes to predicting AQI, and this further highlights the importance of data-driven solutions in tracking the environment. Emissions from vehicles are a primary source of air pollution and contribute to significant AQI values. Machine learning methods provide an efficient way to predict AQI, supporting environmental evaluation and decision-making for preventing air pollution impact.
Keywords: Air pollution, Vehicle emission, Machine Learning, Air Quality Index

INTRODUCTION

Now everyone wants clean and good air, but people do not get air, its pollution. In today's world air pollution is an important environmental problem. All the countries are facing the same problem of air pollution and all of them are trying to reduce it in one way or the other. They have also implemented various rules and practices to reduce air pollution, but it is still a big problem and not a 100% solution [1,12].

They emit large amounts of pollutants and particulate matter (PM), both of which pose a high risk to health because of exposure. As per the figures from the world health organization, The high mortality rate from air pollution is because of maximum population of world lives in areas where air quality in very high mode [1].

As we know, one of the main sources of air and particulate matter (PM) pollution is road transport. This sector mainly uses oil, diesel, kerosene and methanol kind of fossil fuels [4]. According to a German study, the transport sector of our economy emits almost a third portion of lead, nitrogen oxides and all volatile organic compounds.

In Germany, 3% of car trips were discovered to be under 500m. 33.33% of car journeys are found to be not over 3 km, and 50% of car journeys are not over 5 km. These visits normally last a very short time, about 15 minutes. That implies that private cars are highly inefficient when it comes to using energy. When used for short trips, car engines overheat and lose efficiency through the incomplete combustion of fuel, greatly influencing the composition and quantity of organic emissions of the vehicles [2, 21].

Therefore, especially in many cities, there is the problem of traffic and traffic congestion. As a result, when cars are stuck in traffic, more fuel is used, and emissions rise much faster than they actually do. Furthermore, Bull (1991) showed that older cars contribute disproportionately to air pollution [3].

AIR QUALITY INDEX [AQI]

Government agencies employ several measures to give information regarding present air pollution levels. A rise in the Air Quality Index (AQI) signifies a greater threat to public health in everyday life. Nations have their own air quality indices and national air quality standards; for example, Canada's is known as the Air Quality Health Index. The air quality index is split into various classes, with a corresponding color code, description, and prescribed health guidelines. Several factors are input to calculate the AQI, which includes data on traffic as well as industrial discharges [5,16,22]. The National Air Quality Index (AQI) was launched on 17 September 2014 in New Delhi, India under the Swachh Bharat Abhiyan. [Eleven]. National Weather Monitoring Program (NAMP). Continuous monitoring system with real-time data has been set up in Ahmedabad, Kolkata, Mumbai, New Delhi and Pune. Air quality index is categorized into six categories: Good, Satisfactory, moderately polluted, Poor, Very Poor, and Severe. The air quality index suggested will consider eight pollutants PM10, PM2.5, NO2, SO2, CO, O3, NH3, and Pb [13,25,30].

1. Vehicle Emission

Exhaust gases from vehicles emit a large number of harmful substances into the environment. This material is called a pollutant. On human health and the environment, it has many adverse effects. Due to the high demands of people, for air pollution in the world, road traffic is the most important source. In urban areas and major cities of India, Air pollution from cars is a major problem. Car pollution causes ailments like headaches, cough and vision problems from time to time [7,26].

1.2 Vehicle Pollutants

Cars produce a lot of pollutants, depending on the quality of the type of vehicle, fuel, maintenance, etc. The main pollutants emitted by vehicle exhaust gases are oxides of nitrogen (NOx), namely carbon monoxide (CO), butadiene (C4H6), particulate matter (PM), airborne toxins, benzene (C6H6), photochemical oxidants, hydrocarbons (HC), aldehydes, polycyclic aromatic hydrocarbons (PAHs), and lead sulfur oxides (SO2) [6,15,17].

1.3 Vehicle Pollution In India

One of the major environmental issues of Air pollution in our country are most of the people are suffering from health problems due to poor air quality. Rapid urbanization in India has led to an increase in the number of cars. That's why the crowd is increasing. Today, the main source are cars of air pollution in Indian cities. The country will need to improve traffic planning and management by, among other things, improving fuel quality, passing necessary legislation and enforcing vehicle emission standards. He is trying to improve the quality of urban air in many ways [9,10,20].

2. Machine Learning Prediction Methods

In this study, machine learning algorithms were utilized to predict future air quality index values. Several techniques, including Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Logistic Regression, Random Forest (RF), and Decision Tree (DT), were applied. To develop classification models, data modeling software and tools such as NAM and Orange were employed. These tools facilitate various tasks, including (1) dataset loading, (2) segmentation of the dataset into 70% for training and 30% for validation, (3) model construction, (4) storing the model on a hard drive, (5) testing and prediction, and (6) displaying performance scores [11,18].

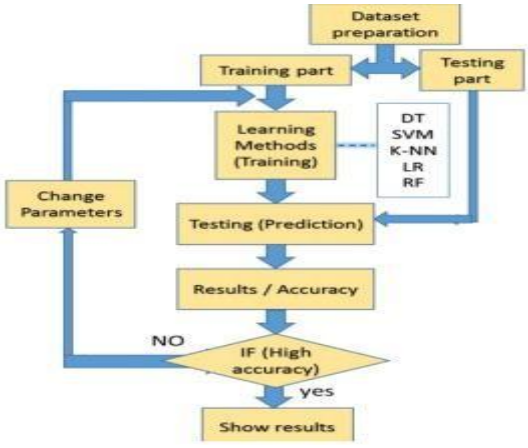

Figure 1 Basic flow of machine learning algorithm [21]

2.1 SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a machine learning algorithm that can be used for classification, regression, and the detection of outliers. It does this by creating a hyperplane that is a separator of points of data with the objective of maximizing separation. Figure 2 illustrates two uses of the SVM technique. Support vectors are the points at the boundary of the region nearest to the hyperplane in Figure 2(a), and the boundary between two distinct classes is determined by the space between these regions. The hyperplane parameter determines the number of classes in the data set, and it is utilized to predict new, unseen observations and to determine the nearest class. In Figure 2(b), the linear model predicts a hyperplane with the maximum margin. The ε-insensitive loss is another very important parameter in this approach [27,28].

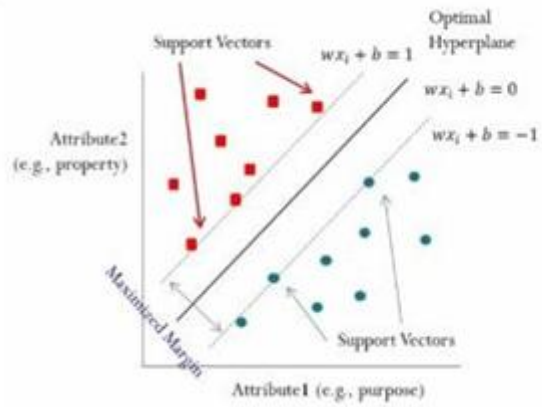

Figure 2 (a) SVM for classification [7]

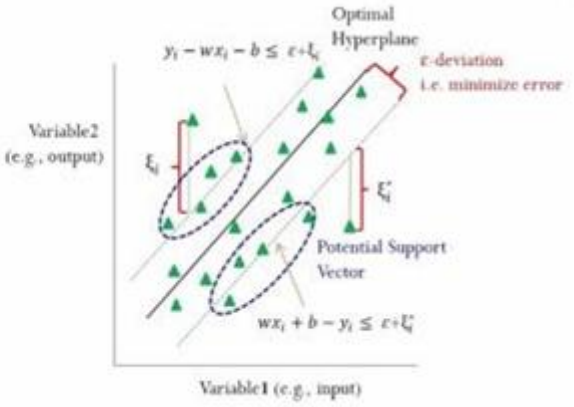

Figure 2 (b) SVM for regression.[7]

2.2 Random Forest

Random Forest is yet another machine learning method applied for prediction of data. It is a supervised learning algorithm and works as an ensemble method by using many decision trees. Random Forest adds randomness in the process of model building by using two major mechanisms: randomly selecting features to distinguish trees and randomly selecting samples to create subsets of training data for every decision tree. At prediction time, every test point traverses all the decision trees in the forest to contribute towards the ultimate result. An example of a Random Forest model with multiple decision trees is shown in Figure 3 [19,20].

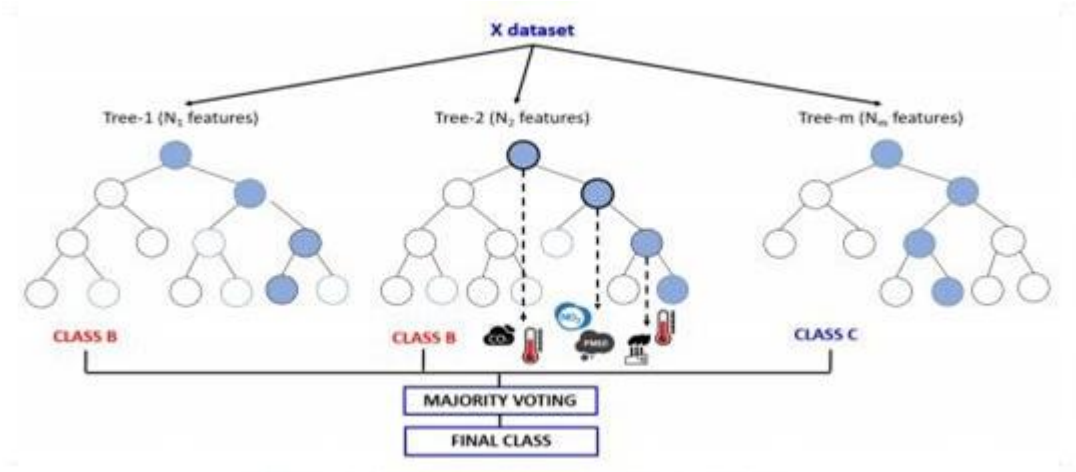

Figure 3 Basic flow of a random forest algorithm [20]

2.3 Decision Tree [Dt]

For predicting, one of the most common and effective used algorithms is Decision tree method. Decision tree is a concept tree model for prediction. Each internal node represents functions that multiply the data using statistics such as profit and information gain. Partitioning the data until it reaches the leaf nodes is an iterative process. For medium, it provides 99.928% accuracy [23].

2.4 Logistic Regression

Logistic Regression is a very popular statistical machine learning classifier. Logistic Regression considers results in terms of non-linear functions, such as the logistic function and sigmoid function. This model has been found to achieve a 91.598% accuracy [23].

2.5 K-Nearest Neighbor (K-Nn)

K-Nearest Neighbors (K-NN) is a lazy learning model that is widely applied to classification and regression. It categorizes data points according to their similarity to other samples in the training set by using distance measures such as Euclidean and Manhattan distances. The most accurate result obtained was 93.886% when applying K-NN with k=6 and 99.714% when applying a numerical distance measure [23,24].

3. Impiementation Methodology

This study collected the main pollutant emissions from any transport sector. India monitors six pollutants (O3, PM2.5, PM10, CO, SO2 and NO2). AQI is calculated manually using equations (1) and (2), where O3, PM2.5 and PM10 are required to determine AQI.

$$AQI = \begin{cases} max\{I_{O3}, I_{PM2.5}, I_{PM10}, I_{CO}, I_{SO2}, I_{NO2}\}, I_{O3}, I_{PM2.5}, I_{PM10} \neq \varnothing \\ \varnothing, otherwise \end{cases} \quad (1)$$

Pollutant concentration ($value_i$) is converted to pollutant index (I_i) by the following formula:

$$I_i = LB_j + \frac{value_i - lb_i}{ub_i - lb_i} \times \left(UB_j - LB_j\right) \quad (2)$$

Here i = O3, NO2, SO2, PM2.5, CO, PM10, and j from specific pollutants is the concentration using the different categories like wise "good", "moderate", "unhealthy" according to the air quality index level. It also includes certain groups such as unhealthy, very unhealthy, and dangerous.

4. CONCLUSION

Demand and base on that production in the transportation sector are increasing rapidly with rapid globalization and urbanization. Therefore, To the development of more vehicles and the deterioration of air quality around the world. By analyzing the data and based on predictions made using various machine learning methods, we can show that vehicle emissions from the transport sector contribute significantly to air pollution. In recent time, most people in every country suffer from air pollution and poor air quality, which can cause lung cancer, asthma, etc. lead to the spread of diseases. New innovative technologies, alternative fuels and government policies can reduce pollution from vehicle emissions.

If implemented correctly, we can improve air quality and achieve a more sustainable environment. Drive as little as possible, government regulations and policies, invest in zero-emission vehicles, have proper pollution control technology, there are many ways to control vehicle emissions, such as scrapping old cars. Also, if people change their minds, we can make great progress. In this way, we can reduce the vehicle pollution and get clean air and good air quality index.

REFERENCES

1. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016).
2. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, (2001).
3. World Health Assembly. Health, Environment and Climate Change: Report by the Director-General; World Health Assembly: Geneva, Switzerland, 2018; 7p
4. Goral Godhani, Maulik Dhamecha, "A Study on Movie Recommendation System Using Parallel MapReduce Technology", IJEDR (2017)
5. Maulik Dhamecha, Dr. Amit Ganatra, Dr. C. K. Bhensadadiya, "Comprehensive Study of Hierarchical Clustering Algorithm and Comparison with Different Clustering Algorithms", CiiT (2011)
6. Dhaval Chandarana, Maulik Dhamecha, "A Survey for Different Approaches of Outlier Detection in Data Mining", IEEE (2015)

7. J. Patel et al., "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," Expert Systems with Applications, vol. 42, no. 1, pp. 259–268, (2015).
8. Vehicle emissions and their effects on natural environments by Journal of the Ghana Institutionof Engineers Vol. 4, No.1, 2006, pp 35–41 35 (Old-3)
9. Bull, K. R. (1991): The critical loads/levels approach to gaseous emission control.Environmental Pollution, Vol. 69, pp. 105 – 123. (Old-4)
10. Maulik Dhamecha, Dr. Tejas Patalia, "Fundamental Survey of Map Reduce in Bigdata with Hadoop Environment", Spinger – CCIS (2018)
11. Chen, T., & Guestrin, C., "XGBoost: A scalable tree boosting system". In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794) (2016).
12. Krishna Parmar, Nivid Limbasiya, Maulik Dhamecha, "Feature based Composite Approach for Sarcasm Detection using MapReduce", IEEE (2018)
13. Khan, A., & Anwar, F., "Stock market prediction using machine learning algorithms". International Journal of Computer Applications, 177(38), 6–10 (2020).
14. Fischer, T., & Krauss, C., "Deep learning with long short- term memory networks for financial market predictions". European Journal of Operational Research, 270(2), 654–669 (2018).
15. Maulik Dhamecha, Dr. Tejas Patalia, "MapReduce Foundation of Big data with Hadoop environment", ELSEVIER - SSRN (2018)
16. Md Shahjada Alam, Dr. Arif Khan, "The Impact Study of Vehicular Pollution on Environment", IJSART (2020) (Old-7)
17. Muhammad, I.; Yan, Z. Supervised Machine Learning Approaches: A Survey. Ictact J. SoftComput. 2015, 5, 946–952. [CrossRef] (Old-8)
18. Maulik Dhamecha, Krupa Dobaria, Dr. Tejas Patalia, "A Survey on Recommendation System for Bigdata using MapReduce Technology", IEEE (2019)
19. Nelson, D. M., Pereira, A. C. M., & de Oliveira, R. A., "Stock market's price movement prediction with LSTM neural networks". International Joint Conference on Neural Networks (IJCNN), IEEE (2017).
20. Maulik Dhamecha, Dr. Tejas Patalia, "Comparative study of Dynamic Load Balancing algorithm in large scale data (Big data)", IJAST (2020)
21. Urvi Darji, Shruti Oza, Kruti P. Thakor "Design Of PFD With Charge Pump For DPLL In 0.18µm CMOS Technology", NCAET (2012)
22. Arumugam, T., Arun, R., Anitha, R., Swerna, P. L., Aruna, R., & Kadiresan, V. (2024). Advancing and Methodizing Artificial Intelligence (AI) and Socially Responsible Efforts in Real Estate Marketing. In S. Singh, S. Rajest, S. Hadoussa, A. Obaid, & R. Regin (Eds.), Data-Driven Intelligent Business Sustainability (pp. 48-59). IGI Global. https://doi.org/10.4018/979-8-3693-0049-7.ch004
23. P. Kranthi, Dr Kiran Kumar Thoti, Dr Shilpa Bhakar, Dr R. Arun, & Dr Biswo Ranjan Mishra. (2024). The Relationship Between Sustainable Human Resource Management and Green Human Resource Management- A Case of Medical Sector in Hyderabad, India. South Eastern European Journal of Public Health, 401–410. https://doi.org/10.70135/seejph.vi.1190
24. Arun, D. R., Srivastava, D. S., Sireesha, rajeyyagari, & Thilagaraj.A , D. (2024). Examining the Varied Culinary Practices and Nutritional Patterns: A Comprehensive Analysis of Eating Habits Amongst Student Populations in Educational Settings. South Eastern European Journal of Public Health, 392–400. https://doi.org/10.70135/seejph.vi.1189
25. Maulik Dhamecha, Dr. Tejas Patalia, "Scheduling issue for Dynamic Load Balancing of mapreduce in large scale data (Big data)", Journal of Xidian University (2020)
26. Maulik Dhamecha, "Improve K-Mean Clustering Algorithm in Large-Scale Data for Accuracy Improvement", AISC (2021)
27. Maulik Dhamecha, "An Efficient Comparison between Structured Analysis and Object-Oriented Analysis", IJSRCSEIT (2021)
28. Dash, S., Sahoo, J., & Mishra, B. B., "A hybrid CNN-LSTM model for stock price prediction". Procedia Computer Science, 167, 2314–2323 (2021).
29. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É., "Scikit-learn: Machine learning in Python". Journal of Machine Learning Research, 12, 2825–2830 (2011).
30. Maulik Dhamecha, "Plant Recommendation System with the Use of Weather Forecasting", CEUR (2021)
31. Awad, M.; Khanna, R. "Support Vector Regression. In Efficient Learning Machines", Apress: Berkeley, CA, USA (2015) (Old-9)
32. D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, "Logistic regression". Springer,2002 (Old-11)
33. H. Wang, C. Tseng, and T. Hsieh, "Developing an indoor airquality index system based on the health risk assessment," Proceedings of indoor air, (2008) (Old-12)
34. Khalid M.O.Nahar1, Mohammad Ashraf Ottom2, Fayha Alshibli3 and Mohammed M. Abu Shquier, "Air quality index using machine learning – a Jordan case study", COMPUSOFT (2020) (Old-13)
35. Rama Lakshmi "India launches its own Air Quality Index. Can itsnumbers be trusted?", (2014) (Old-14)
36. "Ambient Air Quality Monitoring Stations" (http://www.cpcb.nic.in/) (Old-15)
37. Central Pollution Control Board in India (http://www.cpcb.nic.in/) (Old-16)
38. Arun R, and Bhuvaneswari R (2019). Buying behavior of meet's consumption relates to food safety from north and south part of the Coimbatore City. International Journal of Recent Technology and Engineering, 7, 429-433. https://www.ijrte.org/wp-content/uploads/papers/v7i5s/ES2177017519.pdf
39. Arun, R., et al. "From Data to Decisions on Artificial Intelligence's Influence on Digital Marketing Research." Optimizing Intelligent Systems for Cross-Industry Application, edited by S. Suman Rajest, et al., IGI Global, 2024, pp. 1-18. https://doi.org/10.4018/979-8-3693-8659-0.ch001
40. Arumugam, T., Arun, R., Natarajan, S., Thoti, K. K., Shanthi, P., & Kommuri, U. K. (2024). Unlocking the Power of Artificial Intelligence and Machine Learning in Transforming Marketing as We Know It. In S. Singh, S. Rajest, S. Hadoussa,

A. Obaid, & R. Regin (Eds.), Data-Driven Intelligent Business Sustainability (pp. 60-74). IGI Global. https://doi.org/10.4018/979-8-3693-0049-7.ch005
41. Maulik Dhamecha, "Optimize Ant Colony algorithm with relation of efficiency and performance for the Traveling Salesman Problem", GRADIVA (2021)
42. Brownlee, J., "Machine Learning Mastery With Python". Machine Learning Mastery (2019).
43. Rundo, F., Trenta, F., & Battiato, S., "Machine learning for financial forecasting: Theory and applications". Future Internet, 11(10), 1–23 (2019).
44. Maulik Dhamecha, "Fundamentals of ANN and basic algorithm of Neural Network in data analytics", MOENIA (2021)
45. Maulik Dhamecha, "Study of search engines with the paradigm of machine learning", SAMRIDDHI (2022)
46. Raju, P., et al. "Next-Generation Management on Exploring AI-Driven Decision Support in Business." Optimizing Intelligent Systems for Cross-Industry Application, edited by S. Suman Rajest, et al., IGI Global, 2024, pp. 61-78. https://doi.org/10.4018/979-8-3693-8659-0.ch004
47. R. Arun, M. Umamaheswari, K. Premalatha, M. V. Kumar, A. Stella and S. Pl, "Stress Management Through Workplace Associations with Productivity and Mood: The Impact of Learning Experience Based on Hybrid RF-GA-DNN Approach," 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC), Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/ICECCC61767.2024.10593908
48. Chandar, K. A., & Anbazhagan, N., "Ensemble learning approach for stock market forecasting using bagging and boosting". Journal of Intelligent & Fuzzy Systems, 43(1), 1275–1287 (2022).
49. Urvi Darji, Shruti Oza, Kruti P. Thakor "Low power and low jitter phase frequency detector for dpll using 0.18µm cmos technology", ICIKR-ETS (2012)
50. Maulik Dhamecha, "Fundamental study of chatbot with the use of artificial intelligence", SODHASAMHITA (2022)
51. Maulik Dhamecha, "Literature survey for outlier detection methods in data analytics", TANZ Research Journal (2023)
52. Maulik Dhamecha, "Literature survey of basic methodologies in cloud computing", SAMRIDDHI (2022)
53. Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K., "Deep learning for stock prediction using numerical and textual information". IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS) (2016).
54. Barunik, J., & Křehlík, T., "Measuring the frequency dynamics of financial connectedness and systemic risk". Journal of Financial Econometrics, 16(2), 271–296 (2018).
55. Kingma, D. P., & Ba, J., "Adam: A Method for Stochastic Optimization". International Conference on Learning Representations (ICLR) (2015).
56. Goodfellow, I., Bengio, Y., & Courville, A., Deep Learning. MIT Press (2016).