

Feature Selection Optimization for Air Quality Prediction Using Machine Learning in Urban Environments

Nader Behdad, Weiguo Gee, Wang Zhang

1. Electrical and Computer Engineering, The Polytechnic University of the Philippines, Manila, 1016, Philippines
2. School of Computer System, Hebei University of Engineering, Handan, Hebei, 056038, China
3. School of Earth and Space Sciences, Peking University, Beijing, 100871, China

Abstract:

Accurate prediction of air quality is essential for public health management and environmental sustainability in urban areas. This study proposes an optimized machine learning framework for air quality prediction, emphasizing feature selection to enhance model accuracy and computational efficiency. Multiple urban datasets comprising meteorological parameters, vehicular emissions, industrial activity, and pollutant concentrations are preprocessed and subjected to advanced feature selection techniques, including Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and metaheuristic-based methods such as Genetic Algorithms (GA). These selected features are then used to train machine learning models including Random Forests, Support Vector Machines (SVM), and Gradient Boosting. Results indicate a significant improvement in prediction accuracy and processing time, supporting proactive environmental management and policy development in smart city contexts.

Keywords:

Air quality prediction, feature selection, machine learning, urban environments, smart cities, environmental monitoring.

REQUEST FOR FULL TEXT

REFERENCES

- [1] El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., Eid, M. M., & Hussein, S. E. (2020). Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE access*, 8, 179317-179335.
- [2] El-Kenawy, E. S. M., Eid, M. M., Saber, M., & Ibrahim, A. (2020). MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection. *IEEE Access*, 8, 107635-107649.
- [3] El-Kenawy, E. S., & Eid, M. (2020). Hybrid gray wolf and particle swarm optimization for feature selection. *Int. J. Innov. Comput. Inf. Control*, 16(3), 831-844.
- [4] El-Kenawy, E. S. M., Khodadadi, N., Mirjalili, S., Abdelhamid, A. A., Eid, M. M., & Ibrahim, A. (2024). Greylag goose optimization: nature-inspired optimization algorithm. *Expert Systems with Applications*, 238, 122147.
- [5] El-Kenawy, E. S. M., Mirjalili, S., Ibrahim, A., Alrahmawy, M., El-Said, M., Zaki, R. M., & Eid, M. M. (2021). Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification. *Ieee Access*, 9, 36019-36037.
- [6] Abdelhamid, A. A., El-Kenawy, E. S. M., Khodadadi, N., Mirjalili, S., Khafaga, D. S., Alharbi, A. H., ... & Saber, M. (2022). Classification of monkeypox images based on transfer learning and the Al-Biruni Earth Radius Optimization algorithm. *Mathematics*, 10(19), 3614.
- [7] Ibrahim, A., Mirjalili, S., El-Said, M., Ghoneim, S. S., Al-Harhi, M. M., Ibrahim, T. F., & El-Kenawy, E. S. M. (2021). Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm. *IEEE Access*, 9, 125787-125804.
- [8] El-Kenawy, E. S. M., Mirjalili, S., Alassery, F., Zhang, Y. D., Eid, M. M., El-Mashad, S. Y., ... & Abdelhamid, A. A. (2022). Novel meta-heuristic algorithm for feature selection, unconstrained functions and engineering problems. *IEEE Access*, 10, 40536-40555.
- [9] Abdelhamid, A. A., El-Kenawy, E. S. M., Alotaibi, B., Amer, G. M., Abdelkader, M. Y., Ibrahim, A., & Eid, M. M. (2022). Robust speech emotion recognition using CNN+ LSTM based on stochastic fractal search optimization algorithm. *Ieee Access*, 10, 49265-49284.
- [10] Abdollahzadeh, B., Khodadadi, N., Barshandeh, S., Trojovský, P., Gharehchopogh, F. S., El-kenawy, E. S. M., ... & Mirjalili, S. (2024). Puma optimizer (PO): a novel metaheuristic optimization algorithm and its application in machine learning. *Cluster Computing*, 27(4), 5235-5283.
- [11] Eid, M. M., El-kenawy, E. S. M., & Ibrahim, A. (2021, March). A binary sine cosine-modified whale optimization algorithm for feature selection. In *2021 National Computing Colleges Conference (NCCC)* (pp. 1-6). IEEE.
- [12] El-Kenawy, E. S. M., Mirjalili, S., Abdelhamid, A. A., Ibrahim, A., Khodadadi, N., & Eid, M. M. (2022). Meta-heuristic optimization and keystroke dynamics for authentication of smartphone users. *Mathematics*, 10(16), 2912.
- [13] Abdelhamid, A. A., Towfek, S. K., Khodadadi, N., Alhussan, A. A., Khafaga, D. S., Eid, M. M., & Ibrahim, A. (2023). Waterwheel plant algorithm: a novel metaheuristic optimization method. *Processes*, 11(5), 1502.
- [14] Alhussan, A. A., Abdelhamid, A. A., El-Kenawy, E. S. M., Ibrahim, A., Eid, M. M., Khafaga, D. S., & Ahmed, A. E. (2023). A binary waterwheel plant optimization algorithm for feature selection. *IEEE Access*, 11, 94227-94251.
- [15] Hassib, E. M., El-Desouky, A. I., Labib, L. M., & El-Kenawy, E. S. M. (2020). WOA+ BRNN: An imbalanced big data classification framework using Whale optimization and deep neural network. *soft computing*, 24(8), 5573-5592.
- [16] El-Kenawy, E. S. M., Abdelhamid, A. A., Ibrahim, A., Mirjalili, S., Khodadadi, N., Alduailij, M. A., ... & Khafaga, D. S. (2023). Al-Biruni Earth Radius (BER) Metaheuristic Search Optimization Algorithm. *Comput. Syst. Sci. Eng.*, 45(2), 1917-1934.
- [17] Alharbi, A. H., Towfek, S. K., Abdelhamid, A. A., Ibrahim, A., Eid, M. M., & Khafaga, D. S. & Saber, M. (2023). Diagnosis of Monkeypox Disease Using Transfer Learning and Binary Advanced Dipper Throated Optimization Algorithm. *Biomimetics*, 8(3), 313.
- [18] El-Kenawy, E. S. M., Mirjalili, S., Khodadadi, N., Abdelhamid, A. A., Eid, M. M., El-Said, M., & Ibrahim, A. (2023). Feature selection in wind speed forecasting systems based on meta-heuristic optimization. *Plos one*, 18(2), e0278491.
- [19] Khodadadi, N., Abualigah, L., El-Kenawy, E. S. M., Snasel, V., & Mirjalili, S. (2022). An archive-based multi-objective arithmetic optimization algorithm for solving industrial engineering problems. *IEEE Access*, 10, 106673-106698.
- [20] Eid, M. M., El-Kenawy, E. S. M., Khodadadi, N., Mirjalili, S., Khodadadi, E., Abotaleb, M., ... & Khafaga, D. S. (2022). Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of monkeypox cases. *Mathematics*, 10(20), 3845.
- [21] Khodadadi, N., Khodadadi, E., Al-Tashi, Q., El-Kenawy, E. S. M., Abualigah, L., Abdulkadir, S. J., ... & Mirjalili, S. (2023). BAOA: binary arithmetic optimization algorithm with K-nearest neighbor classifier for feature selection. *IEEE Access*, 11, 94094-94115.
- [22] Salamai, A. A., El-kenawy, E. S. M., & Abdelhameed, I. (2021). Dynamic voting classifier for risk identification in supply chain 4.0. *Computers, Materials & Continua*, 69(3).
- [23] Djaafari, A., Ibrahim, A., Bailek, N., Bouchouicha, K., Hassan, M. A., Kuriqi, A., ... & El-Kenawy, E. S. M. (2022). Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions. *Energy Reports*, 8, 15548-15562.