

<https://doi.org/10.7250/CONNECT.2026.112>

# EXPLAINABLE MACHINE-LEARNING MODELOPTIMIZED USING FORWARD FEATURE SELECTION ALGORITHM FOR PREDICTION OF GROUND OZONE CONCENTRATION

Ibrahim KHALIL UMAR<sup>1\*</sup>, Samir BASHIR<sup>2</sup>

<sup>1</sup> *Department of Civil and Engineering, Federal University of Technology, Babura Jigawa, Nigeria*

<sup>2</sup> *Department of Civil and Engineering Technology, Kano State Polytechnic, Kano*

\* *Corresponding author. Email address: ikumar.civil@futb.edu.ng*

**Abstract** – An accurate model for prediction of air quality parameters such as ground ozone (o<sub>3</sub>) serves as policy and decision-making tool for providing healthy and friendly environment. in this study, six machine-learning models (neural network (nn), ensemble, kernel regression, regression tress, support vector regression (svr), and multilinear regression (mlr)) were developed for prediction of ground ozone concentration using 35 065 hourly data from January 2013 to february 2017. The selection of relevant input variables for the models was using forward feature selection algorithm. the forward feature selection reveals rain, day and wind speed as the least important variable for the prediction of ozone concentration. the models were evaluated using mean absolute error (mae), mean square error (mse), root mean square error (rmse) and nash sutcliffe efficiency (nse) in both training and testing stage. the nn model outperformed ensemble, kernel, tree, svr and mlr by 0.73 %, 4.74 %, 6.47 %, 11.41 % and 25.17 % respectively in the testing stage. copland algorithm indicates the nn model as the overall best model considering all evaluation metrics. the shapley analysis indicates temperature, nitrogen oxide and hour of the day as the major factors contributing to ground ozone concentration.

**Keywords** – *Ground ozone, neural network, machine learning, forward feature selection*