EPH53

Analysis of the Contribution of Air Pollution and Noise Levels to Predicting Autonomic Nervous System Disorders: An Interpretable Machine Learning Approach

Hyeonjung Park[1,2], Hae Sun Suh[1,2,3*]

1 Department of Regulatory Science, Graduate School, Kyung Hee University, Seoul, Republic of Korea
2 Institute of Regulatory Innovation through Science (IRIS), Kyung Hee University, Seoul, Republic of Korea
3 College of Pharmacy, Kyung Hee University, Seoul, Republic of Korea *Corresponding author

PEBD Pharmaceutical Economics Big Data Analysis and Policy Lab

Hyeonjeong.park@khu.ac.kr

Background

- **Autonomic nervous system (ANS):** Sustained sympathetic activation may disrupt autonomic balance, while persistent reductions in parasympathetic activity have been associated with elevated risks of disease and mortality[1,2].
- **Environmental stressors:** Noise and air pollution are growing public health concerns[3]. Noise activates the sympathetic nervous system, triggering stress hormones and the renin-angiotensin system[3]. Inhaled particles and gases cause inflammation and stress, affecting autonomic activity through nerve signals[4,5].
- **Objectives:** We aimed to assess the contribution of environmental stressors to ANS disorder prediction using machine learning (ML) models and Shapley Additive Explanations (SHAP).

Methods

Data Processing

UK biobank whole population (n=502,367)

Data set	▪ Included participants with inpatient records after 2011 ▪ Baseline: first assessment visits (2006–2010) ▪ Outcome: ANS-related ICD-10 codes
Pre-processing	▪ Ecological variable removal: >30% missing ▪ Others removal: >50% missing ▪ Missing data imputation: KNN method
Split data	Training 70% / Validation 30%
Sampling	Random undersampling (case : control = 20:80)
Standardization	Standardization based on mean (SD) / Validation 30%
Feature selection	Feature selection Backward elimination ($p < 0.05$) / Validation 30%

Statistical analysis

Data set

- **Database:** UK Biobank
- **Variables**
  - **Baseline characteristics**: sex, age, ethnicity, income, education, TDI
  - **Covariates**: smoking, alcohol use status, nutritional intake, BMI, physical activity score(minutes/week)
  - **Envirionmental exposure:** surrounding natural environment, noise levels, air pollution (including PM, NO, and NO_2), and traffic density.
  - **ANS-related ICD10 codes:** G900, G901, G902, G903, G904, G908, G909, I951

Model develop

- **ML models:** Logistic Regression, Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Stacking Ensemble, Fully Connected Neural Network (FCNN)

Model explanation

- **Performance**: accuracy, precision, recall, F1-score, ROC-AUC
- **SHAP plot**: feature importance visualization

Results

Baseline characteristics

- After the data processing, ANS incident people was 4,910 and not incident people was 12,831. (20:80)

Characteristics	Developed ANS disease	No ANS disease
Age, mean (SD)	59.23 (7.95)	64.09 (6.28)
Male gender, n (%)	8,912 (45.4%)	2,807 (57.2%)
Ethnicity, n (%)		
White	17,892 (91.1%)	4,516 (92.0%)
Mixed	788 (4.0%)	182 (3.7%)
Others	935 (4.9%)	204 (4.3%)
Education score, mean (SD)	15.77 (15.65)	18.42 (17.85)
Income score, mean (SD)	0.12 (0.09)	0.13 (0.11)
Townsend deprivation index, mean (SD)	-1.27 (3.11)	-0.78 (3.30)
Smoking status		
Never	10,327 (52.6%)	2,252 (45.9%)
Previous	7,120 (36.3%)	2,013 (41.0%)
Current	2,088 (10.6%)	594 (12.1%)
Alcohol drinker status		
Never	882 (4.5%)	316 (6.4%)
Previous	737 (3.8%)	334 (6.8%)
Current	17,994 (91.6%)	4,247 (86.5%)
BMI, mean (SD)	27.66 (4.91)	28.34 (5.28)

Abb. ANS; Autonomic nervous system.

SHAP plot

- SHAP values were calculated for the best-performing stacking model.
- Greater distance from the coast and higher PM2.5 concentrations positively contributed to the prediction of ANS disease occurrence.
- Lower noise levels contributed more strongly to the prediction of ANS disease occurrence.

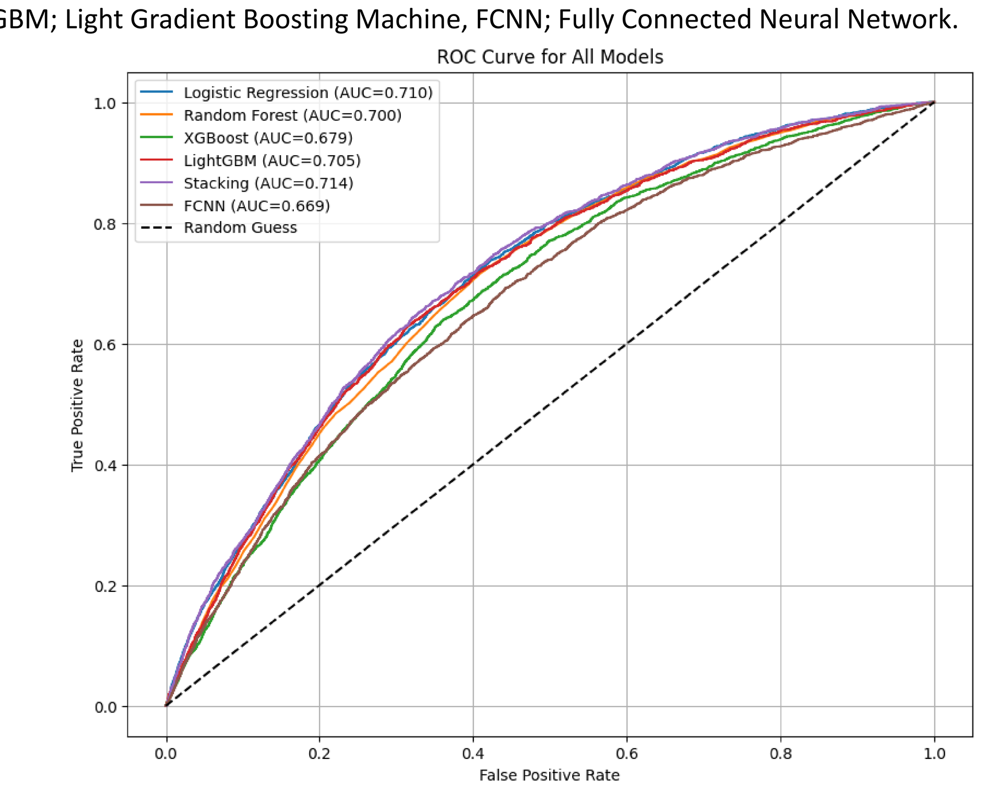
Model performance

Model	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	0.6994	0.0370	0.6100	0.0697	0.7105
Random Forest	0.6967	0.0372	0.6204	0.0702	0.7127
XGBoost	0.6944	0.0367	0.6166	0.0693	0.7145
LightGBM	0.6959	0.0369	0.6171	0.0697	0.7146
Stacking	**0.7009**	0.0374	0.6152	0.0706	**0.7153**
FCNN	0.6847	0.0360	0.6233	0.0680	0.7125

Abb. XGBoost; Extreme Gradient Boosting, LightGBM; Light Gradient Boosting Machine, FCNN; Fully Connected Neural Network.

- The best-performing model was the stacking model, achieving an accuracy of 0.705 and an AUC of 0.714.

LOC curve



Discussion

Summary

- Predictive models for ANS disease were developed using ML.
- Contributions of environmental exposures were evaluated by SHAP, with minimal impact from air pollution and noise.

Limitations

- Environmental exposure was assessed only once at baseline (2006–2010), not longitudinally.
- Residential changes during follow-up and individual outdoor activity levels were not considered.

Study implications

- This study evaluated environmental factors for ANS disorders using a large-scale population database with interpretable ML approaches.
- The application of SHAP enables the generation of scientific evidence to support public health improvements.
- Future studies should incorporate longitudinal exposure data and account for individual activity patterns to more accurately assess environmental influences.

Conclusions

- Environmental factors, including air pollution and noise, showed limited contribution to ANS disorder prediction in this study.
- Future studies addressing limitations in exposure assessment and study design are needed to validate these findings.

Acknowledgment

- This work was a grant (21153MFDS601) from the Ministry of Food and Drug Safety in 2025.
- This research was supported by a grant (RS-2024-00331719) from the Ministry of Food and Drug Safety (MFDS).
- This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00345981).
- This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. RS-2024-00509257, Global AI Frontier Lab)

References

1. Thayer, Julian F., and Richard D. Lane. "The role of vagal function in the risk for cardiovascular disease and mortality." Biological Psychology 74.2 (2007): 224–242.
2. Thayer, Julian F., et al. "A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health." Neuroscience & Biobehavioral Reviews 36.2 (2012): 747–756..
3. Münzel, Thomas, et al. "Environmental risk factors and cardiovascular diseases: A comprehensive expert review." Cardiovascular Research 118.14 (2022): 2880–2902.
4. Perez, Christina M., Mehdi S. Hazari, and Aimen K. Farraj. "Role of autonomic reflex arcs in cardiovascular responses to air pollution exposure." Cardiovascular Toxicology 15 (2015): 69–78.
5. Niemann, Bernd, et al. "Oxidative stress and cardiovascular risk: Obesity, diabetes, smoking, and pollution: Part 3 of a 3-part series." Journal of the American College of Cardiology 70.2 (2017): 230–251.