

NDnano Summer Undergraduate Research 2026 Project Summary

1. Student name & home university: Matthew Kim - University of Notre Dame
2. ND faculty name & department: Prof. Nosang V. Myung, Department of Chemical and Biomolecular Engineering
3. Summer project title: Chemiresistive-Based Electronic Nose (E-Nose) System with Multi-Stage Machine Learning for Objective Halitosis Diagnosis
4. Briefly describe new skills you acquired during your summer research:
Training and validating machine learning and deep learning models, soldering, electrospinning, 3D printing and modeling
5. Briefly share a practical application/end use of your research:
The E-Nose system has various applications such as identifying chemical warfare agents in the environment, detecting avian influenza, and diagnosing halitosis.
6. 50- to 75-word abstract of your project:

Halitosis affects many people and signals poor oral hygiene or periodontal disease, yet clinical diagnosis relies on subjective organoleptic evaluation. We developed an electronic nose using 16 chemiresistive sensors on tunable microhotplates to detect three volatile sulfur compounds and four breath biomarkers. A multi-stage machine learning pipeline classifies analytes (Random Forest, KNN, Neural Networks) and predicts concentration via regression. Ongoing work focuses on improving ML performance and validating the system with healthy and infected artificial breath samples.

7. References for papers, posters, or presentations of your research:
[1] S. H. Yoo, H. S. Jung, W. S. Sohn, B. H. Kim, B. H. Ku, Y. S. Kim, S. W. Park, and K. B. Hahm, "Volatile sulfur compounds as a predictor for esophagogastrroduodenal mucosal injury," Gut Liver, vol. 2, no. 2, pp. 113–118, 2008.
[2] Agency for Toxic Substances and Disease Registry, ToxGuide for Hydrogen Sulfide (H₂S), CAS# 7783-06-4. Atlanta, GA, USA: U.S. Dept. of Health and Human Services, Public Health Service, Dec. 2016.
[3] Agency for Toxic Substances and Disease Registry, "Methyl mercaptan: Medical management guidelines (MMG)," Toxic Substances Portal, CAS# 74-93-1. [Online]. Available: <https://wwwn.cdc.gov/tsp/MMG/MMGDetails.aspx?mmgid=221&toxid=40>
[4] American Industrial Hygiene Association, Odor Thresholds for Chemicals with Established Health Standards, 2nd ed. Falls Church, VA, USA: AIHA, 2013.

[5] M. Li, C. Wu, Z. Wang, Z. Wu, W. Huang, J. Chen, K. Yu, T. Wen, H. Yin, and Z. Wang, "Enhanced gas classification in electronic nose systems using an SMOTE-augmented machine learning framework," *Sensors*, vol. 26, no. 2, p. 714, 2026.

[6] D. R. Wijaya, F. Afianti, A. Arifianto, D. Rahmawati, and V. S. Kodogiannis, "Ensemble machine learning approach for electronic nose signal processing," *Sens. Bio-Sens. Res.*, vol. 36, p. 100495, 2022.

One-page project summary that describes problem, project goal, and your research activities / results:

Halitosis (bad breath) affects a substantial portion of the population and can indicate poor oral hygiene, gingivitis, or periodontitis. The clinical standard for diagnosis, organoleptic evaluation, relies on a trained examiner directly smelling a patient's breath, making it subjective and error-prone. To address this, we developed an e-nose system for faster, more objective, and more accurate diagnosis. The system consists of an array of 16 chemiresistive sensors, each with a unique sensing material dropcasted onto an individual microhotplate; varying microhotplate temperature tunes each sensor's response to different analytes. Sensor chips were exposed to three volatile sulfur compounds (VSCs): methyl mercaptan, hydrogen sulfide, and dimethyl sulfide, key indicators of halitosis, along with four common breath metabolites (acetone, isoprene, phenol, and d-limonene). Nine key features were extracted per sensor (16), chip (5), concentration (7), analyte (7), and operating temperature (8), feeding into a multi-stage machine learning pipeline that first classifies the analyte and then estimates its concentration through regression.

For analyte classification, Extra Trees achieved the best accuracy at 88%, with Random Forest a close second (Figure 1); decision-tree-based ensemble methods outperformed other approaches overall. For concentration regression, an MLP regressor reached up to 80% accuracy (R^2) for acetone, but performed poorly for other analytes, with dimethyl sulfide regression R^2 close to 0. We then moved to a simpler least-squares regression approach, fitting each analyte individually using its single best-performing sensor channel and operating temperature, which produced substantially better and more consistent results across all analytes.

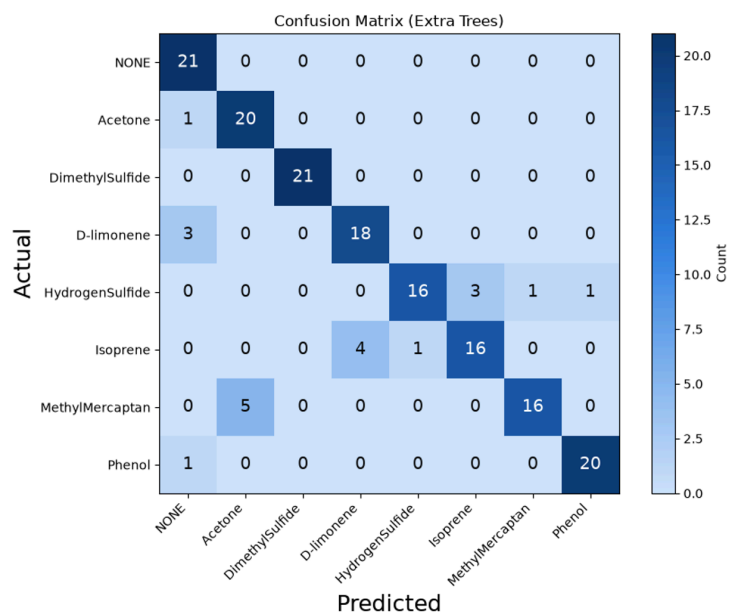


Figure 1. Extra Trees confusion matrix for analyte classification

This least-squares approach was evaluated using leave-one-out cross-validation, comparing each sensor's predicted analyte concentration (log10 scale) against the actual known concentration for that sample. Plotting predicted versus actual values, with a dashed diagonal line representing a perfect prediction (predicted = actual), showed that points clustering tightly around this line indicate accurate, reliable predictions. Hydrogen sulfide and methyl mercaptan both reached an R^2 of 0.847 using their best single sensor channel (C08 at 450°C and 350°C, respectively), with predicted points falling close to the ideal line across the full concentration range tested. Dimethyl sulfide performed noticeably worse, with an R^2 of only 0.654 on its best channel (C08 at 450°C): while predictions were reasonably accurate at low concentrations, they diverged more from the actual values at higher concentrations, showing the sensor array has room to improve for quantifying this particular compound. Overall, these results indicate that our single-sensor least-squares method substantially outperforms the more complex MLP regressor for most analytes, though further refinement is still needed to improve consistency for compounds like dimethyl sulfide.