

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 42, Pages 30-38

ICBASET 2026: International Conference on Basic Sciences, Engineering and Technology

Performance Evaluation of Machine Learning Models for Gas Classification Under Sensor Drift on ESP32-S3

Gonca Uslu-Ozkucuk
OSTIM Technical University

Abstract: Machine learning has become an effective approach for electronic nose systems used in gas classification applications. However, most previous studies evaluate classification performance using randomly partitioned datasets, which often fail to represent the long-term sensor drift encountered in practical deployments. Furthermore, embedded implementation of different machine learning models on resource-constrained microcontrollers has received limited attention. In this study, the Gas Sensor Array Drift Dataset from the UCI Machine Learning Repository was used to investigate both classification performance and embedded deployment efficiency. Four supervised machine learning algorithms, namely Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were evaluated under both random stratified and chronological batch-based data partitioning strategies. The chronological evaluation was designed to simulate realistic sensor drift conditions by separating the training and testing data according to acquisition time. Experimental results showed that all models achieved approximately 99% accuracy under random evaluation, whereas chronological testing produced substantially lower accuracies, highlighting the impact of sensor drift on real-world performance. Among the evaluated models, MLP achieved the highest chronological classification accuracy while demonstrating superior robustness against distribution shifts. To investigate practical applicability, all trained models were deployed on an ESP32-S3 microcontroller. Their average inference latency, flash memory usage, RAM consumption, and total firmware size were experimentally measured and compared. Among the evaluated models, the MLP achieved the highest chronological classification accuracy while also exhibiting the lowest inference latency and smallest firmware footprint on the ESP32-S3 platform.

Keywords: Electronic nose, TinyML, ESP32-S3, Gas classification, Sensor drift

Introduction

Electronic nose (e-nose) systems have become an important technology for gas identification and environmental monitoring in applications such as food quality assessment, industrial safety, environmental monitoring, and healthcare (Han et al., 2025; Ijaz et al., 2025; Illahi et al., 2021; Ivanov et al., 2025; Rabehi et al., 2024). These systems typically employ arrays of metal oxide semiconductor (MOS) gas sensors together with machine learning algorithms to recognize complex gas patterns (Ijaz et al., 2025; Vanaraj et al., 2025; Wijaya et al., 2022; Yi et al., 2021). Compared with conventional analytical instruments, electronic noses provide a low-cost, portable, and real-time solution for gas classification (Anwar et al., 2023; Baruah & Mazumder, 2025; Ivanov et al., 2025; Yang et al., 2023).

The performance of machine learning-based electronic nose systems, however, is strongly affected by sensor drift, which refers to the gradual change in sensor characteristics over time due to aging, contamination, and environmental conditions (Cinar et al., 2025; Dennler et al., 2022; Ziyatdinov et al., 2010). As sensor responses change, the statistical distribution of the collected data also changes, causing a significant degradation in classification accuracy when models are deployed in long-term applications (Cheng et al., 2020; R. Yi et al.,

2021). Therefore, developing machine learning models that are robust to sensor drift has become an important research topic (Chang et al., 2023; Leon-Medina et al., 2020).

Previous studies have investigated various machine learning algorithms for gas classification using the Gas Sensor Array Drift Dataset, including Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost) (Cinar et al., 2025; Dennler et al., 2022; Leon-Medina et al., 2020; R. Yi et al., 2021). Most of these studies have primarily focused on improving classification accuracy in an offline computing environment. In addition, many published results have relied on random train-test splitting strategies, where training and testing samples originate from the same data distribution (Dong et al., 2021; Gkogkidis et al., 2022; Shamim, 2022; Z. Yi et al., 2022). Although this approach often produces classification accuracies exceeding 99%, it does not accurately represent the performance of electronic nose systems operating under long-term sensor drift conditions (Dennler et al., 2022; Padilla et al., 2010).

Another limitation of the existing literature is that relatively few studies evaluate machine learning models on resource-constrained embedded platforms (Rahmawan et al., 2025; Shamim, 2022). TinyML has recently emerged as a promising approach for executing machine learning algorithms directly on low-power microcontrollers, enabling intelligent edge devices without continuous cloud connectivity (Farhat et al., 2025; Han et al., 2025). While TensorFlow Lite Micro has facilitated the deployment of neural network models on embedded systems, comparative studies involving both deep learning and classical machine learning models under identical embedded conditions remain limited (Farhat et al., 2025; Ivanov et al., 2025; Nivethika et al., 2026). Furthermore, most previous studies report only offline classification performance without experimentally evaluating memory usage and inference latency on actual hardware (Shamim, 2022).

To address these limitations, this study compares four supervised machine learning models—Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and XGBoost—for gas classification under sensor drift conditions using the Gas Sensor Array Drift Dataset (Farbiz et al., 2025; Ivanov et al., 2025; Wang et al., 2023). Two evaluation protocols are considered: a conventional random stratified split and a chronological split that preserves the temporal order of measurements (Ansari et al., 2026; Kumar et al., 2026; Padilla et al., 2010; Yi et al., 2023). In addition to offline performance evaluation, all trained models are deployed on an ESP32-S3 DevKitC-1 N16R8 development board (Farhat et al., 2025; Nivethika et al., 2026; Rahmawan et al., 2025). The embedded implementations are experimentally compared in terms of classification consistency, firmware size, memory usage, and inference latency.

The main contributions of this study can be summarized as follows; the classification performance of four widely used machine learning models is systematically compared under both random and chronological evaluation protocols. The influence of sensor drift on model generalization is quantitatively analyzed using chronological validation. All trained models are successfully deployed on the ESP32-S3 platform, enabling a direct comparison of their embedded implementation characteristics. The proposed study jointly evaluates classification accuracy, memory footprint, and inference latency, providing practical guidance for selecting suitable machine learning models for TinyML-based electronic nose applications.

Overall, the proposed methodology provides a comprehensive evaluation framework that considers both classification performance under realistic sensor drift conditions and the computational constraints of embedded hardware, thereby contributing to the development of efficient intelligent gas sensing systems.

Method

This study used the Gas Sensor Array Drift Dataset from the UCI Machine Learning Repository to investigate the impact of sensor drift on gas classification performance (VERGARA, 2012). The dataset consists of 13,910 measurements belonging to six gas classes. Each sample contains 128 features, extracted from eight response characteristics of 16 metal oxide (MOS) gas sensors. The measurements are organized into 10 chronological batches, enabling the evaluation of long-term sensor drift effects (Vergara et al., 2012).

The dataset was processed in the Python environment by separating the feature vectors and class labels to construct the data matrix. Since the dataset contained no missing or erroneous records, no additional data cleaning procedures were required. Feature values were normalized using Min–Max scaling. The normalization parameters were calculated exclusively from the training data and subsequently applied to the validation and test sets to prevent data leakage.

Two evaluation protocols were employed. In the random stratified split, the dataset was divided into 70% training, 15% validation, and 15% testing while preserving the class distribution. In the chronological split, Batches 1–7 were used for training, Batch 8 for validation, and Batches 9–10 for testing. This protocol enabled the evaluation of model generalization under realistic sensor drift conditions. Figure 1 illustrates the overall experimental workflow adopted in this study.

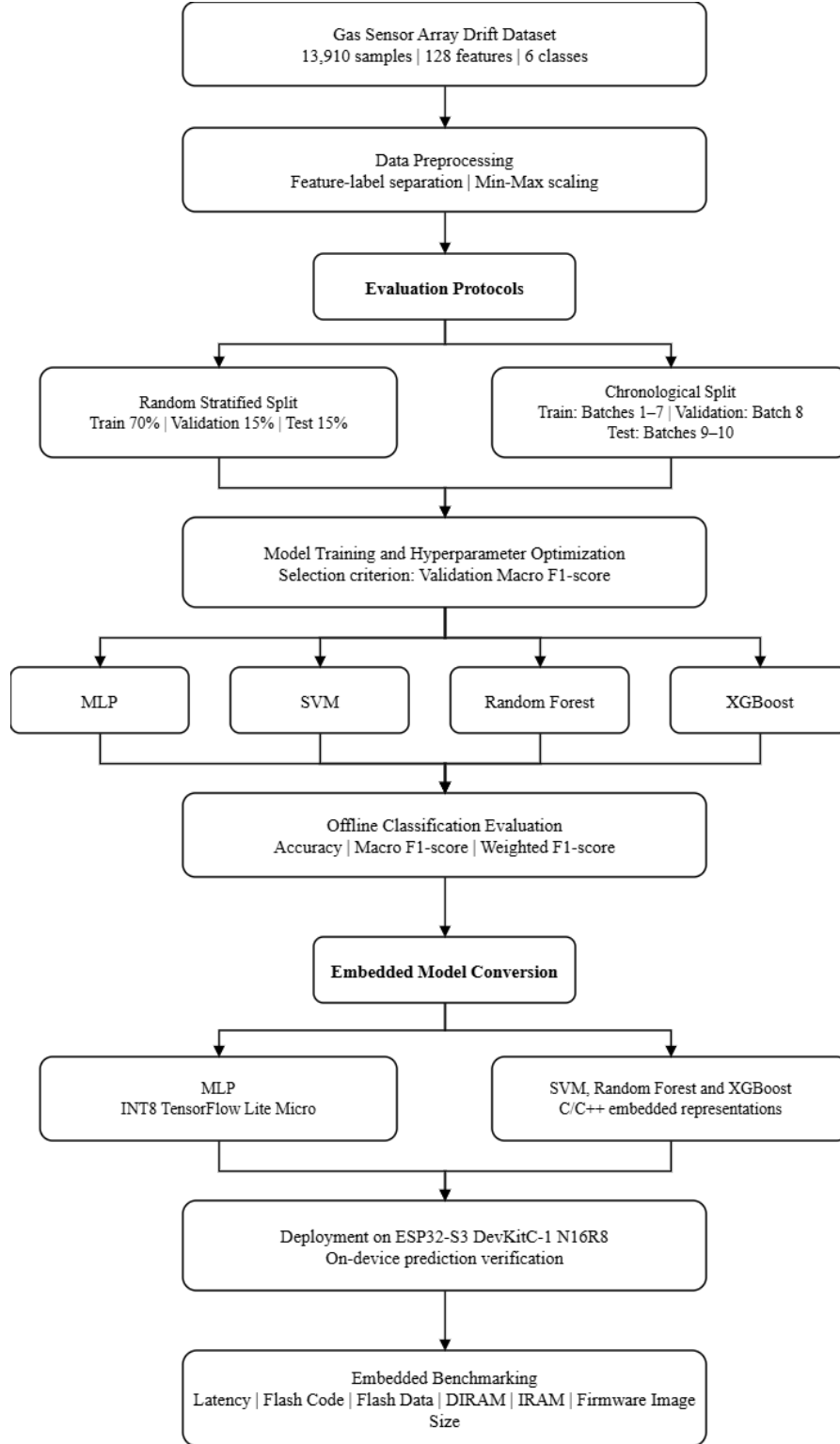


Figure 1. The overall experimental workflow of this study

Four supervised machine learning models were investigated: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The MLP model was

implemented using TensorFlow (Abadi et al., 2016), the SVM and Random Forest models were developed using Scikit-learn (Pedregosa FABIANPEDREGOSA et al., 2011), and the XGBoost model (Chen & Guestrin, 2016) was implemented using the XGBoost library. Hyperparameters were optimized using the training and validation datasets, and the final model configurations were selected according to the highest validation Macro F1-score.

All trained models were deployed on an ESP32-S3 DevKitC-1 N16R8 development board. The MLP model was converted to an INT8 TensorFlow Lite Micro model (David et al., 2021) using post-training quantization. The SVM, Random Forest, and XGBoost models were automatically converted into C/C++ header files and integrated into the ESP32-S3 firmware. The embedded applications were developed using the ESP-IDF framework, and all inference operations were executed locally on the microcontroller without cloud connectivity. Model performance was evaluated using Accuracy, Macro F1-score, and Weighted F1-score. For the embedded implementation, average inference latency, Flash Code, Flash Data, DIRAM, IRAM, and total firmware image size were measured to compare the computational efficiency and memory requirements of the deployed models.

Results and Discussion

A chronological data splitting approach was used to evaluate the performance of machine learning models under real-world operating conditions. In this method, Batches 1–7 were used for training, Batch 8 for validation, and Batches 9 and 10 for testing data. Thus, the temporal data distribution change caused by sensor drift was directly included in the evaluation process (Vergara et al., 2012).

Table 1. Chronological test performance on batches 9 and 10

Model	Accuracy (%)	Macro-F1 (%)	Weighted-F1 (%)	Batch 9 acc. (%)	Batch 10 acc. (%)	Prediction/sample (ms)
MLP	69.34	69.51	69.55	76.38	68.42	0.0349
SVM	63.88	65.22	65.20	75.96	62.31	0.0476
Random Forest	59.63	59.43	59.45	76.17	57.47	0.0148
XGBoost	58.18	58.36	58.36	71.49	56.44	0.0027

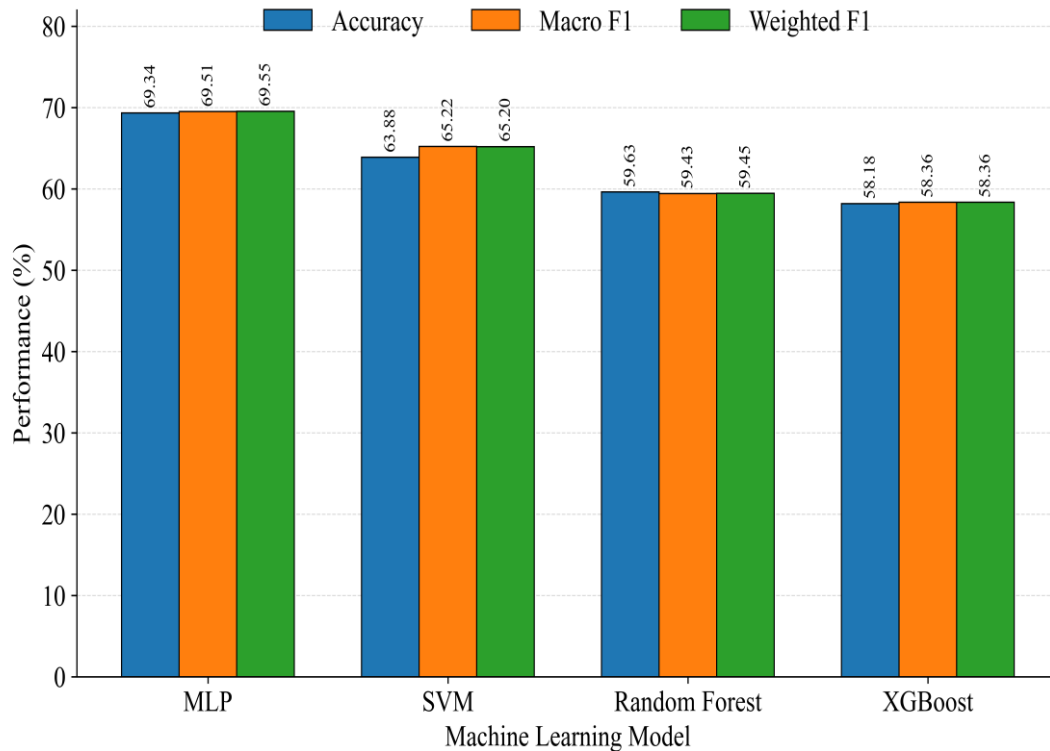


Figure 2. Accuracy comparison of models in chronological evaluation

Table 1 summarizes the numerical performance of the evaluated models, while Figure 2 provides a visual comparison of their classification accuracy. Among the four models, the MLP model achieved the highest

classification accuracy. With 69.34% accuracy, 69.51% Macro F1, and 69.55% Weighted F1, the MLP model outperformed the other models. The SVM model ranked second with 63.88% accuracy, while the Random Forest and XGBoost models achieved 59.63% and 58.18% accuracy, respectively. Furthermore, the MLP model maintained the highest accuracy values in both Batch 9 and Batch 10, demonstrating more stable performance against sensor drift.

These results show that the MLP model has a higher generalization ability than other methods against data distributions that change over time due to sensor drift. In particular, the fact that the accuracy and F1 scores are quite close to each other shows that the model exhibits a balanced classification performance not only in dominant classes but in all gas classes. In contrast, tree-based models, although they successfully learned the training data, struggled to adapt to temporal changes. The most commonly used evaluation method in machine learning studies is the random data splitting approach. However, in datasets with sensor drift, this approach does not fully reflect real-world performance because training and test samples do not come from the same distribution.

Table 2. Random stratified test performance

Model	Accuracy (%)	Macro-F1 (%)	Weighted-F1 (%)	Prediction/sample (ms)
Random Forest	99.47	99.44	99.47	0.0335
XGBoost	99.33	99.29	99.33	0.0022
MLP	99.23	99.18	99.23	0.0373
SVM	99.19	99.16	99.18	0.0773

Table 3. Comparison of random stratified and chronological evaluation protocols

Model	Random acc. (%)	Chronological acc. (%)	Accuracy gap (pp)	Random macro-F1 (%)	Chronological macro-F1 (%)	Macro-F1 gap (pp)
Random Forest	99.47	59.63	39.84	99.44	59.43	40.01
XGBoost	99.33	58.18	41.15	99.29	58.36	40.92
MLP	99.23	69.34	29.90	99.18	69.51	29.67
SVM	99.19	63.88	35.30	99.16	65.22	33.94

When Tables 2 and 3 are examined together, it is seen that while all models in the random data splitting approach achieve approximately 99% accuracy, the accuracy values drop to the 58–69% range under chronological evaluation. This demonstrates that random evaluation substantially overestimates the real-world performance of gas classification models in the presence of sensor drift. The accuracy loss in the MLP model is approximately 30 percentage points, while in the Random Forest and XGBoost models this loss reaches approximately 40 percentage points. The highest performance decrease was observed in the XGBoost model.

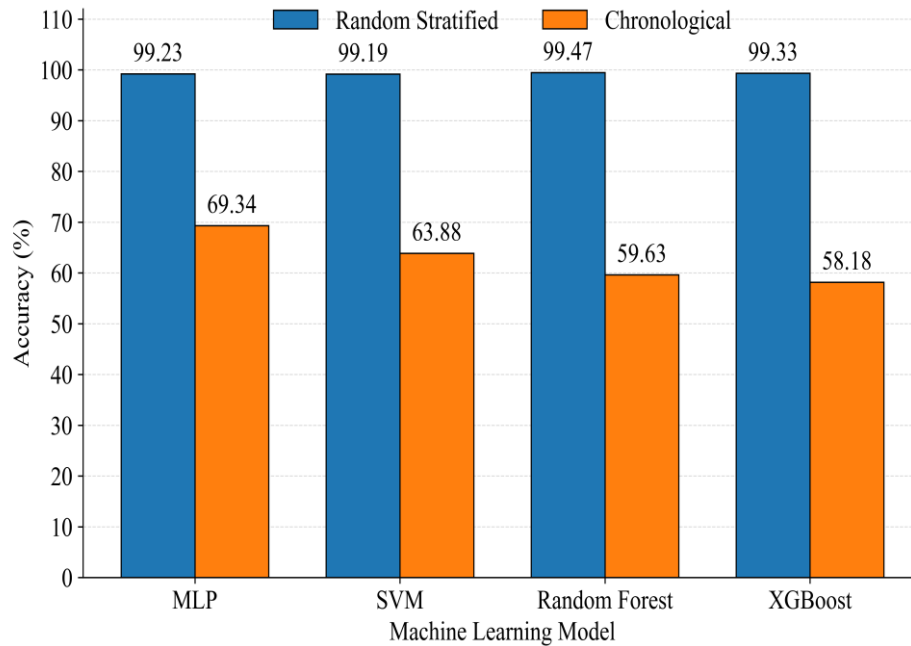


Figure 3. Comparison of random stratified and chronological classification accuracy.

In Figure 3 comparison of random stratified and chronological classification accuracy is showed. These results reveal that high accuracy values reported using only the random data splitting approach in datasets with sensor drift may overestimate performance under real operating conditions. Therefore, the chronological evaluation approach offers a more realistic method for assessing the generalization ability of models in long-term sensor applications such as electronic noses. In the second phase of the study, the four trained machine learning models were transferred to the ESP32-S3 DevKitC-1 N16R8 development board and run on real hardware. This allowed for the evaluation not only of offline classification success but also of the models' applicability on a microcontroller with limited resources.

Table 4. Comparison of memory usage and inference time of machine learning models on ESP32-S3

Model	Flash Code (B)	Flash Data (B)	DIRAM (B)	Firmware Size (B)	Avg. Latency (μ s)
MLP (INT8)	98,156	65,404	49,990	222,530	401.64
Random Forest	902,680	38,180	45,126	999,814	5,117
XGBoost	62,780	370,484	45,126	492,218	12,600
SVM (RBF)	63,476	429,916	48,070	552,346	16,995

Table 4 shows the memory usage and inference time results obtained on the ESP32-S3. The MLP model was the fastest with an average inference time of 401.64 μ s. The average inference times for the Random Forest, XGBoost, and SVM models were measured as 5117 μ s, 12600 μ s, and 16995 μ s, respectively. Accordingly, the MLP model operates with approximately 13 times lower latency than Random Forest, 31 times lower latency than XGBoost, and 42 times lower latency than the SVM model. Furthermore, it was confirmed that the predictions obtained on the ESP32-S3 for all models were fully consistent with the results obtained in the Python environment.

Significant differences were also observed in terms of memory usage. The MLP model had the lowest memory requirement with a total firmware size of approximately 222.5 kB, while the Random Forest model reached the largest firmware size with approximately 999.8 kB. The firmware sizes of the XGBoost and SVM models were measured as approximately 492.2 kB and 552.3 kB, respectively. When the distribution of resource usage is examined, it is seen that Flash Code usage increased significantly in the Random Forest model due to the direct compilation of decision trees as program code. In contrast, Flash Data usage was higher in the SVM and XGBoost models because a large portion of the model parameters were stored in fixed data space. The MLP model, however, provided the most balanced resource usage in terms of both program code and data memory.

When these results are evaluated together, it is seen that the MLP model not only achieved the highest classification success under chronological evaluation, but also provided the lowest memory requirement and the shortest inference time on the embedded system. In contrast, while the Random Forest model offers relatively low latency, it suffers from a memory disadvantage due to its high firmware size. The SVM model achieved the second highest accuracy in chronological evaluation but had the longest inference time. The XGBoost model yielded moderate results in terms of both accuracy and hardware performance.

Overall, the proposed MLP-based TinyML implementation provides the best trade-off between classification performance, memory footprint, and inference latency, making it a promising solution for resource-constrained electronic nose and IoT applications. These findings demonstrate that deep learning-based models offer significant advantages in electronic nose applications with sensor drift, not only in terms of offline classification success but also in terms of real-time embedded systems applications (Fonollosa et al., 2015; Vergara et al., 2012).

Conclusion

In this study, the performance of four machine learning models (MLP, SVM, Random Forest, and XGBoost) was investigated for gas classification under sensor drift conditions using the Gas Sensor Array Drift Dataset. Unlike conventional studies that rely only on random data splitting, both random stratified and chronological evaluation protocols were employed to assess the effect of temporal distribution changes caused by sensor drift.

The experimental results demonstrated that random stratified evaluation produced classification accuracies of approximately 99% for all models, whereas chronological evaluation reduced the accuracy to 58–69%, revealing a substantial degradation in performance due to sensor drift. Among the evaluated models, the MLP achieved

the highest chronological performance with 69.34% Accuracy, 69.51% Macro F1-score, and 69.55% Weighted F1-score, indicating superior generalization capability under long-term sensor drift.

In addition to offline evaluation, all trained models were successfully deployed on an ESP32-S3 DevKitC-1 N16R8 development board, enabling a direct comparison of their embedded implementation characteristics. Experimental measurements showed that the MLP model required the smallest firmware size and achieved the shortest inference latency (401.64 μ s), while maintaining complete agreement between Python and ESP32 predictions. In contrast, Random Forest exhibited the largest firmware footprint, whereas SVM and XGBoost required considerably longer inference times due to their computational complexity and memory organization.

Overall, the results demonstrate that the MLP model provides the most favorable balance between classification accuracy, memory consumption, and inference latency, making it the most suitable candidate for TinyML-based electronic nose applications on resource-constrained microcontrollers. The proposed evaluation methodology also highlights the importance of chronological validation for obtaining realistic performance estimates in sensor drift scenarios, where conventional random evaluation may significantly overestimate model performance.

Future work will focus on evaluating additional lightweight neural network architectures, investigating online adaptation techniques for long-term sensor drift compensation, and validating the proposed approach using real-time gas measurements acquired directly from an ESP32-based electronic nose prototype.

Recommendations

The proposed methodology can be extended by evaluating additional lightweight deep learning architectures specifically designed for TinyML applications. Future studies may also investigate online or continual learning techniques to compensate for long-term sensor drift without retraining the model offline. Furthermore, the proposed approach should be validated using real-time measurements collected from a physical electronic nose system instead of publicly available datasets. Finally, evaluating power consumption and energy efficiency on battery-powered embedded platforms would provide further insight into the practical applicability of the proposed system.

Scientific Ethics Declaration

* The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM Journal belongs to the author.

Conflict of Interest

* The author declares that there is no conflict of interest.

Funding

* The author received no financial support for the research, authorship, and publication of this article.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Istanbul/Türkiye on September 17-18, 2026.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2016). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*. Retrieved from <http://arxiv.org/abs/1603.04467>
- Ansari, G., Singh, R., Kumar, S., & Kumar, R. (2026). Drift-resilient hybrid feature learning framework for accurate mixed industrial gas classification using chemical sensor arrays. *IEEE Sensors Journal*, 26(10), 15007–15019.
- Anwar, H., Anwar, T., & Murtaza, S. (2023). Review on food quality assessment using machine learning and electronic nose system. *Biosensors and Bioelectronics: X*, 14, 100365.
- Baruah, S., & Mazumder, D. H. (2025). A review on application of machine learning techniques coupled with E-nose in healthcare, agriculture, and allied domains. *IEEE Transactions on Instrumentation and Measurement*, 74.
- Chang, I. S., Byun, S. W., Lim, T. B., & Park, G. M. (2023). A study of drift effect in a popular metal oxide sensor and gas recognition using public gas datasets. *IEEE Access*, 11, 26383–26392.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Cheng, Y. C., Chou, T. I., Chiu, S. W., & Tang, K. T. (2020). A concentration-based drift calibration transfer learning method for gas sensor array data. *IEEE Sensors Letters*, 4(10).
- Cinar, R. F., Yuksek, G., Lale, T., Ekinci, S., Izci, D., & Ma'arif, A. (2025). SHAP-based framework for temporal detection of sensor drift in gas sensor arrays. *Journal of Robotics and Control (JRC)*, 6(5), 2592–2601.
- David, R., Duke, J., Jain, A., Reddi, V. J., Jeffries, N., Li, J., ... Warden, P. (2021). *TensorFlow Lite Micro: Embedded machine learning on TinyML systems*. Retrieved from <http://arxiv.org/abs/2010.08678>
- Dennler, N., Rastogi, S., Fonollosa, J., van Schaik, A., & Schmuker, M. (2022). Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks. *Sensors and Actuators B: Chemical*, 361.
- Dong, X., Han, S., Wang, A., & Shang, K. (2021). Online inertial machine learning for sensor array long-term drift compensation. *Chemosensors*, 9(12).
- Farbiz, F., Aggarwal, S., Maszczyk, T. K., Habibullah, M. S., & Hamadicharef, B. (2025). Reliability-improved machine learning model using knowledge-embedded learning approach for smart manufacturing. *Journal of Intelligent Manufacturing*, 36(7), 4941–4962.
- Farhat, F., El Hamdi, R., & Dahman, H. (2025). Edge embedded deep learning intelligent electronic nose for real-time gas detection. In *2025 IEEE International Multi-Conference on Smart Systems and Green Process (IMC-SSGP 2025)*.
- Fonollosa, J., Rodríguez-Luján, I., & Huerta, R. (2015). Chemical gas sensor array dataset. *Data in Brief*, 3, 85–89.
- Gkogkidis, A., Tsoukas, V., Papafotikas, S., Boumpa, E., & Kakarountas, A. (2022). A TinyML-based system for gas leakage detection. In *2022 11th International Conference on Modern Circuits and Systems Technologies (MOCASST 2022)*.
- Han, J., Li, H., Cheng, J., Ma, X., & Fu, Y. (2025). Advances in metal oxide semiconductor gas sensor arrays based on machine learning algorithms. *Journal of Materials Chemistry C*, 13(9), 4285–4303.
- Ijaz, U., Ali, M., Ahmad, I., Hamza, S. A., & Kim, H. D. (2025). A comprehensive review of electronic nose systems: Design, sensors, and future directions. *Chemical Engineering Journal*, 524.
- Illahi, A. A. C., Dadios, E. P., Bandala, A. A., & Vicerra, R. R. P. (2021). Electronic nose technology and application: A review. In *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM 2021)* (pp. 1–5).
- Ivanov, S., Wilk-Jakubowski, J. Ł., Ciopiński, L., Pawlik, Ł., Wilk-Jakubowski, G., & Mihalev, G. (2025). Modern trends in the application of electronic nose systems: A review. *Applied Sciences*, 15(19).
- Kumar, G., Terauds, M., Ajayakumar Raji, A., Semenako, J., Smolaninova, V., Sics, P. E., & Malayidinja Poikayil Thankappan, A. K. (2026). Design of a lightweight Edge-AI system for predictive maintenance on ESP32-S3. *Applied Sciences*, 16(11).
- Leon-Medina, J. X., Pineda-Munoz, W. A., & Burgos, D. A. T. (2020). Joint distribution adaptation for drift correction in electronic nose type sensor arrays. *IEEE Access*, 8, 134413–134421.
- Nivethika, G., Pragiya, V., Rohitha, E., Padmaja, Y., & Mari Rajan, S. (2026). CookSafe AI: An Edge-AI framework for intelligent gas knob monitoring and autonomous safety actuation using ESP32 with TinyML. In *2026 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI 2026)*.
- Padilla, M., Perera, A., Montoliu, I., Chaudry, A., Persaud, K., & Marco, S. (2010). Drift compensation of gas sensor array data by orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, 100(1), 28–35.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rabehi, A., Helal, H., Zappa, D., & Comini, E. (2024). Advancements and prospects of electronic nose in various applications: A comprehensive review. *Applied Sciences*, 14(11).
- Rahmawan, J., Yudhana, A., & Risky, J. M. (2025). Temporal gated neural model for odor classification in TinyML devices. In *ICoAIT 2025—1st International Conference on Artificial Intelligence Technology: Artificial Intelligence: Driving Prosperity and Sustainability in the Modern World* (pp. 149–154).
- Shamim, M. Z. M. (2022). TinyML model for classifying hazardous volatile organic compounds using low-power embedded edge sensors: Perfecting Factory 5.0 using Edge AI. *IEEE Sensors Letters*, 6(9).
- Vanaraj, R., I. P., B., Mayakrishnan, G., Kim, I. S., & Kim, S. C. (2025). A systematic review of the applications of electronic nose and electronic tongue in food quality assessment and safety. *Chemosensors*, 13(5), 1–23.
- Vergara, A. (2012). *Gas sensor array drift dataset* [Data set]. UCI Machine Learning Repository.
- Vergara, A., Vembu, S., Ayhan, T., Ryan, M. A., & Homer, M. L. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical*, 167.
- Wang, X., Zhou, Y., Zhao, Z., Feng, X., Wang, Z., & Jiao, M. (2023). Advanced algorithms for low dimensional metal oxides-based electronic nose application: A review. *Crystals*, 13(4).
- Wijaya, D. R., Afianti, F., Arifianto, A., Rahmawati, D., & Kodogiannis, V. S. (2022). Ensemble machine learning approach for electronic nose signal processing. *Sensing and Bio-Sensing Research*, 36, 100495.
- Yang, X., Li, M., Ji, X., Chang, J., Deng, Z., & Meng, G. (2023). Recognition algorithms in E-nose: A review. *IEEE Sensors Journal*, 23(18), 20460–20472.
- Yi, D., Zhang, L., Wang, Z., Wang, L., Duan, S., & Yan, J. (2023). Robust domain correction latent subspace learning for gas sensor drift compensation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(12), 7668–7680.
- Yi, R., Yan, J., Shi, D., Tian, Y., Chen, F., Wang, Z., & Duan, S. (2021). Improving the performance of drifted/shifted electronic nose systems by cross-domain transfer using common transfer samples. *Sensors and Actuators B: Chemical*, 329.
- Yi, Z., Shang, W., Xu, T., Guo, S., & Wu, X. (2022). Local discriminant subspace learning for gas sensor drift problem. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(1), 247–259.
- Ziyatdinov, A., Marco, S., Chaudry, A., Persaud, K., Caminal, P., & Perera, A. (2010). Drift compensation of gas sensor array data by common principal component analysis. *Sensors and Actuators B: Chemical*, 146(2), 460–465.

Author(s) Information

Gonca Uslu- Ozkucuk

OSTIM Technical University

Department of Electricity and Energy

Ankara, Türkiye

ORCID iD: <https://orcid.org/0000-0002-2523-2516>

*Contact e-mail: gonca.usluozkucuk@ostimteknik.edu.tr

*Corresponding author(s) contact email

To cite this article:

Uslu-Ozkucuk, G. (2026). Performance evaluation of machine learning models for gas classification under sensor drift on ESP32-S3. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)*, 42, 30-38.