Evaluating the Impact of Adding Gaussian Noise as Data Augmentation on Electronic Nose for Black Tea Quality Classification

Istifa Shania Putri
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
23523007@std.stei.itb.ac.id

Nur Ulva Maulidevi
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
ulfa@staff.stei.itb.ac.id

Abstract*— This study aims to evaluate whether the addition of Gaussian noise as data augmentation enhances the learning process or merely introduces irrelevant data ("noise") into the system. The evaluation focuses on how the augmented data affects learning outcomes, particularly in scenarios with increasing noise levels. The study also compares classification performance with and without noise augmentation to assess its impact on model generalization and accuracy. Gaussian noise was applied at six controlled variance levels (0.01 to 0.3), and its influence on classification was analyzed across 1D-CNN and 2D-CNN. Statistical analyses using MANOVA and the Kolmogorov-Smirnov test confirmed that Gaussian noise augmentation preserved the core structure of the original data at lower noise levels while introducing realistic variability. The results show that for 1D-CNN, Zhou's architecture consistently achieved 96% accuracy across all noise levels, indicating robustness to added noise. In contrast, 2D-CNN models, particularly ResNet34 with transfer learning, demonstrated exceptional performance at low noise levels (0.01) with an accuracy of 98.66%, but experienced a gradual performance decline as noise levels increased. This research provides insights into the effectiveness of Gaussian noise augmentation for improving model learning and highlights its limitations at higher noise levels. By comparing results with and without noise augmentation, the study demonstrates the importance of noise calibration to maintain the balance between variability and data integrity in classification tasks.

Keywords—Gaussian noise augmentation, Data Augmentation, Electronic Nose, tea quality classification, 1D-CNN, 2D-CNN

I. Introduction

The quality of tea, characterized by attributes such as appearance, aroma, color, and taste, is a critical determinant of consumer preferences and market value. These attributes are influenced by various chemical compounds in tea, such as polyphenols, catechins, flavonoids, and aromatic substances, which contribute to the distinct flavor and aroma appreciated by consumers [1]. Among these attributes, aroma stands out as a key factor in determining tea quality. Traditionally, tea quality assessment relies on human experts, or tea testers, who evaluate the tea through observation, smelling, and tasting. However, this method is prone to errors due to subjectivity and variability in assessments across different testers. Moreover, such evaluations are resource-intensive, costly, and unsuitable for mass testing. Standardizing descriptive terms for tea flavors also poses challenges, particularly for researchers from diverse regions [2].

Similar to many food products, the evaluation of black tea quality continues to depend heavily on human sensory tests. While practical, these methods face limitations, including inconsistency, lack of standardized measurements, and subjectivity [3]. Alternative approaches using advanced chemical analysis methods, such as High-Performance Liquid Chromatography (HPLC), Gas Chromatography (GC), and Capillary Electrophoresis, offer high accuracy but require expensive equipment and skilled personel [4]. These challenges underscore the need for rapid, cost-effective, and user-friendly solutions to address the limitations of existing methods.

Electronic nose (e-nose) technology has emerged as a promising alternative to standard chemical analysis methods, offering a cost-effective and efficient solution [5]. E-noses mimic the human olfactory system by utilizing an array of sensors to detect volatile organic compounds (VOCs) present in gas samples. These sensor arrays enable the characterization and evaluation of complex samples, providing reliable and reproducible analyses of aroma [6]. espite its potential, traditional e-nose systems rely heavily on manual preprocessing and feature extraction techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). These approaches, while effective, are time-consuming, require domain expertise, and risk losing critical latent information in the process [7].

Deep learning has revolutionized feature extraction and pattern recognition by enabling end-to-end learning. Convolutional Neural Networks (CNNs), in particular, have demonstrated their ability to automatically extract meaningful features directly from raw data, reducing the need for manual intervention [8]. However, the effectiveness of deep learning models depends significantly on the availability of labeled training data, which is often limited in e-nose applications. Data collection for e-noses is time-intensive and expensive, posing a major barrier to improving model generalization.

II. Related works

A. Data Augmentation

In the domain of time series, datasets are often challenging to access, primarily due to privacy concerns, which frequently hinder the availability of sufficient data . In the context of electronic noses, obtaining high-quality datasets involves extensive parameter simulations, specialized expertise to conduct operations, and significant time investment. These factors make acquiring sufficient and high-quality datasets for training machine learning models a considerable challenge.

One effective solution to address this issue is data augmentation. By applying techniques such as adding noise, performing permutations, or generating synthetic data, datasets can be expanded to overcome these limitations.

In [9], where the study investigates multi-sensor data by proposing a fusion method that combines frequency-domain features from different sensor channels. They decided to augmentation data using Generative Adversarial Networks (GAN), where GAN is on of data augmentation technique that generate syntentic data. They higlighted that the application of GANs for data augmentation in multi-sensor settings shows great promise but comes with significant challenges. The quality of synthetic data generated by GANs is inconsistent due to the randomness in input, which can occasionally reduce classifier performance rather than enhance it. Furthermore, increasing the amount of synthetic data does not always lead to better results, as the relationship between data quantity and model performance is not straightforward. The success of GAN-generated data also depends heavily on the thoughtful selection of features and classification models, requiring substantial expertise and experimentation. Ensuring data quality often requires an additional validation dataset, complicating the process further. Finally, enhancing class separability by deriving representative features from sensor data is essential yet difficult, impacting the robustness and effectiveness of GAN-generated data. These challenges underscore the importance of meticulous design and validation in utilizing GANs for multi-sensor data augmentation.

Where in [10], they evaluated various data augmentation methods for time series classificatin with neural networks, including permution tehcniques as data augmentation methods. Permuation Their study specifically notes that permutation does not preserve temproal dependencies, making it less effective for dataset where time dependencies are crucial. It observed significantly accuracy degradation for some models when using permutation, particulary for neural networks like LSTM and ResNet.

On the other hand, in [11] estabilshed study emphasizes the advantages of Gaussian noise as a data augmentation technique in multi-sensor applications, showcasing its ability to enhance model performance by reducing test errors (RMSE) in predictions such as temperature and indoor air quality (IAQ). For instance, applying Gaussian noise lowered the RMSE for the IAQ model from 0.311 (using original data) to 0.203. This approach introduces controlled randomness through Gaussian-distributed noise, improving model generalization to unseen data while preserving the dataset's statistical properties. Moreover, Gaussian noise consistently demonstrated performance improvements across various LSTM-based prediction models, illustrating its capability to enrich and strengthen training datasets. Nonetheless, the study notes challenges, including the possibility of generating unrealistic values (e.g., negative data), which could impact data quality in specific contexts.

B. Evaluating Augmented Data

Evaluating augmented data is a critical step to ensure the effectiveness of data augmentation techniques in enhancing model performance. Evaluating augmented data generated by adding Gaussian noise is essential to ensure that the augmented data effectively mimics the original data while introducing meaningful variability. To evaluate augmented data generated by adding Gaussian noise to electronic nose (e-nose) sensor readings, statistical techniques like Multivariate Analysis of Variance (MANOVA) and the Kolmogorov-Smirnov (KS) test are crucial. MANOVA assesses whether there are significant differences in the aroma profiles captured by the original and augmented data across multiple sensor features. A non-significant p-value (e.g., > 0.05) indicates that the augmented data maintains the underlying patterns and relationships of the original aroma profiles, ensuring consistency in the representation of smell. The KS test further examines each sensor feature independently, comparing the distributions of sensor responses between the original and noise-augmented data. A low KS statistic and non-significant p-value suggest that the distributions of augmented data closely mimic the original sensor responses.

In [12], the study utilized Multivariate Analysis of Variance (MANOVA) to validate the similarity between original and augmented datasets, where augmentation was performed using interpolation techniques. MANOVA confirmed no significant difference (p-value = 0.1083) between the mean vectors of the two groups, ensuring that the augmented data maintained consistency with the original dataset's structure. This validation was crucial for demonstrating the reliability of augmented data in enhancing predictions for Chronic Obstructive Pulmonary Disease (COPD) while preserving data authenticity. Data was collected using a spirometer. While in another studies, in [13], The study highlights the importance of statistical tools like the Kolmogorov-Smirnov (KS) test in assessing the similarity of distributions between different datasets. By comparing sensor response patterns, the KS statistic effectively measured the agreement between augmented and original data, ensuring the augmented data preserved the key characteristics of the original odor profiles. While MANOVA evaluates differences in multivariate mean structures, the KS test provides insights into distributional consistency, making it more suitable for evaluating point-by-point deviations across datasets. These complementary approaches highlight different perspectives in validating the quality of augmented data.

C. CNN-Based Methods for Electronic Nose Classification

One-dimensional Convolutional Neural Networks (1D-CNN) are highly advantageous for multi-sensor classification tasks, particularly in electronic nose applications. These systems depend on time-series data from various sensors detecting chemical or gas compositions, where identifying temporal changes is essential. 1D-CNN is ideal for such scenarios due to its strength in capturing spatial relationships and local dependencies within sequential data, making it adept at extracting critical features from sensor outputs. Unlike traditional techniques, which often falter in identifying intricate temporal patterns, 1D-CNN utilizes convolutional layers to focus on significant patterns while filtering out irrelevant information. Its capability to process data of varying lengths and its efficient feature extraction make it a preferred solution for handling the complexity of multi-sensor datasets with accuracy and speed.

In [14] developed a 1D-CNN architecture with additional filters to classify methane, ethylene, ethanol, and carbon monoxide gases from e-nose time-series sensor data. The architecture focuses on leveraging 1D kernels to preserve the structural integrity of e-nose data by preventing convolution across unrelated sensor responses, ensuring reliable feature extraction. The model consists of multiple convolutional layers with a kernel size of 30×1, starting with Conv1, which outputs 32 channels, followed by a pooling layer with a stride of (2,1). Subsequent layers, Conv2 to Conv9, refine the feature extraction using the same kernel size but reduce the output channels to 16, maintaining efficiency while capturing critical patterns. A final linear layer converts the extracted features into a 5-class output for classification. The authors's architecure are well-suited for e-nose time-series data, strong noise reduction and feature extraction capabilities are offered, making it a promising foundation for black tea classification. However, small dataset challenges and data augmentation techniques, such as adding noise, are not addressed in their study, which are critical in this research.

On the other hand, [15] proposed a 1D-CNN architecture for classifying e-nose data, such as essential oils, coffee, and whiskey, using four stacked convolutional blocks for feature extraction. Each block consists of two convolutional layers (kernel size 128×10, stride 1), two batch normalization layers, and a max-pooling layer, with ReLU activation applied after normalization. To reduce complexity, a Global Average Pooling (GAP) layer is used to convert feature maps into a 128-dimensional embedding vector, which is then mapped to predictions through a fully connected (FC) layer, minimizing parameters while maintaining high performance. The architecture's use of stacked convolutional blocks with batch normalization (BN) inherently makes it more resilient to variations in the input, including noisy data. BN helps stabilize learning by normalizing activations [16]. Which can mitigate the effects of added noise during training. The max-pooling and Global Average Pooling (GAP) layers efficiently down-sample features, which can help filter out minor noise-induced variations while retaining critical patterns. However, it is important to note that this architecture has not yet been evaluated on noise-augmented datasets, leaving its full potential for handling artificially enhanced variability unexplored. Incorporating noise-based augmentation in future evaluations could provide valuable insights into the model's robustness and adaptability to real-world sensor data variations.

Another CNN based that can be used to done classification task of electronic nose response is Two-Dimensional Convolutional Neural Network (2D- CNN), owing to their ability to capture spatial patterns within sensor arrays. By transforming e-nose sensor readings into structured 2D representations, 2D-CNNs can effectively analyze correlations between sensors and identify critical features relevant to classification tasks, such as detecting specific gases or odors. In [17], the study focuses on improving the classification accuracy of an ee-nose using 2D-CNN. It investigates how spatial and temporal information from sensors at different locations affects performance. By transforming sensor data into a 2D representation, the study demonstrates that the proposed 2D-CNN outperforms a 1D-CNN alternative, achieving a test accuracy of 97.8%, which is 7.5% higher than the 1D-CNN. The study primarily evaluated sensor performance without explicitly testing the model's robustness to various levels of noise in the sensor data. Another studies in [18], introduced approach using a multichannel CNN for gas mixture classification. It transforms multivariate time-series data from multiple gas sensors into analogous image matrice. To expand the dataset, they applied data augmentation techniques, such as scaling, to simulate additional data. Scaling modifies the range of sensor values while maintaining their relative proportions, making it useful for simulating sensor sensitivity variations. In contrast, Gaussian noise introduces random variability to the data, mimicking real-world sensor noise and enhancing the model's robustness to environmental fluctuations. Different studies, in [19] focused on utilizing 2D-CNN for classifying gases such as methane (CH4), carbon monoxide (CO), and their mixtures using an enose. To prepare the data for 2D CNN processing, the matrices were converted into grayscale images, where each value represented pixel intensity and captured the sensor's response over time. To compensate for the limited dataset size, the study relied heavily on data augmentation techniques (e.g., translation and cropping), which may not fully replicate real-world variations.

D. Converting data sensor into image

Transforming sensor data into image representations is essential for adapting time-series or multivariate sensor inputs for deep learning models, especially CNNs. This conversion enables the visual depiction of patterns and relationships within the data, facilitating more effective feature extraction by CNNs. In [18], input adaptation is performed by transforming gas sensor data for 2D CNN processing. This involves converting multivariate time-series data into analogue image matrices through a process that includes linear interpolation to address data gaps and normalization to scale the data into the [0,1] range. Another studies, in [19], gas sensor data was organized into a matrix format with rows representing time and columns representing sensors. The data was then normalized, rescaled to a [0,255] range, and represented as grayscale images.

The primary difference between the two methods lies in the preprocessing steps and the preservation of spatial-temporal relationships. While [18] uses linear interpolation to handle missing data points, ensuring smooth transitions and continuity, [19] does not explicitly address data gaps but focuses on a simpler transformation workflow. Additionally, the approach in [18] emphasizes scaling the data into a normalized [0,1] range before further processing, which can enhance numerical stability and learning efficiency. The method in [18] surpasses [19] by effectively handling incomplete datasets through interpolation, enhancing feature quality.

III. METHODOLOGY

A. Dataset

The dataset used in this study originates from previous research by [20] and comprises measurements from an electronic nose (e-nose) on three quality levels of black tea: Quality 1 (Q1), Quality 2 (Q2), and Quality 3 (Q3). The black tea samples were sourced from PT. Tambi, a tea factory located in the highlands of Dieng, Wonosobo, Central Java, which produces various grades of black tea. Quality 1 (Q1), known as Broken Orange Pecco (BOP), is finely ground, dark black tea made from tea buds without stems, intended for export. Quality 2 (Q2), or BP II (Broken Pecco II), is coarser and reddish-black, derived from young tea leaves and stems. Quality 3 (Q3), referred to as Bohea, is light brown, coarse tea made from older tea leaves and stems. The tea qualities were verified by testers from PT. Tambi and labeled accordingly.

The dataset consists of 1,550 labeled samples for each tea quality, totaling three classes. These data were collected using 12 gas sensors: MQ-7, TGS 2600, TGS 813, TGS 825, TGS 2602, TGS 826, TGS 2610, TGS 2611, TGS 832, TGS 2612, TGS 2620, and TGS 822. The process of gathering labeled training data for the e-nose model was labor-intensive and time-consuming, highlighting the challenges in building a dataset that balances model performance, cost, and development time.

B. Data Augmentation Process

Data augmentation was performed by adding controlled Gaussian noise to the original dataset, following the formula [11]:

$$x_{augmented} = x_{original} + \mathcal{N}(\mu, \sigma^2) \tag{1}$$

Where $\mathcal{N}(\mu, \sigma^2)$ represents Gaussian noise with mean $\mu = mean\ data$ and variance σ^2. Six levels of noise variance were applied: 0.01, 0.03, 0.05, 0.1, 0.2, and 0.3, resulting in six augmented data groups. These groups were treated separately during experimentation to assess the impact of noise levels on the classification system. This approach aimed to introduce variability in the data while ensuring that the signal remained within the realistic range of sensor readings, mimicking environmental or instrumental noise in real-world conditions.

Class balance was also checked during this process to ensure that the final dataset had an equal number of samples across all three tea quality classes (Q1, Q2, and Q3). Maintaining class balance was critical to prevent biased model training and to enhance the system's generalization ability.

The augmented dataset was introduced into the classification model to evaluate whether the added noise constitutes entirely meaningless data that disrupts the model's learning process or can be considered meaningful input that enriches the training process. This approach aims to assess whether the noise-enhanced data helps the model learn from diverse variations, making it more capable of distinguishing subtle patterns and adapting to real-world scenarios involving sensor or environmental variability.

C. Data Augmentation Evaluation

The evaluation of the augmented data was conducted using statistical tools and visual analysis to compare the aroma profiles of the augmented data with the original data. A radar plot was used to visualize the sensor responses for different classes, allowing for a qualitative comparison of the overall aroma profiles. Additionally, two primary statistical methods were employed: MANOVA and KS test.

A non-significant p-value (e.g., p>0.05) would suggest that the Gaussian noise augmentation preserved the multivariate relationships present in the original aroma profile data. This ensures that the added noise does not distort the fundamental patterns of the sensor responses. For e-nose datasets, MANOVA is particularly useful as it evaluates consistency across multiple sensor responses simultaneously, making it ideal for assessing the preservation of the aroma profile.

The Kolmogorov-Smirnov (KS) test was used to evaluate the similarity of distributions between the original and augmented datasets for each sensor feature individually. The KS test compares the cumulative distribution functions (CDFs) of two datasets, where the KS statistic is defined as [13]:

$$d_{KS} = \max |F_{(x_i)} - F_{(x_j)}| \tag{2}$$

Where *D* is KS statistic; sup denotes the supremum (the maximum absolute difference); $F_{(x_i)}$ and $F_{(x_j)}$ are CDF of data. This formula calculates the maximum distance between the two CDFs, providing a quantitative measure of how similar the distributions of the original and augmented data are. A smaller *D* value indicates greater similarity, and a non-significant p-value suggests that the augmented data retains the characteristics of the original dataset.

To complement the statistical tests, radar plots were generated to provide a visual representation of the aroma profiles for each class. These plots displayed the maximum responses of each sensors for each class, allowing for an intuitive comparison between the original and augmented data. Consistency in radar plot shapes and ranges further validated the similarity of the augmented data to the original profiles.

D. Convert data sensor into image

The sensor data, originally in tabular form, was converted into a 2D representation suitable as input, specifically a 2D-CNN. This step ensures compatibility with the requirements of image-based models, enabling effective feature extraction and classification. The conversion process was adapted from the methodology proposed by [18], with modifications to better suit the characteristics of the current dataset and experimental goals. In this study, we introduced changes to the transformation process to better capture the overall distribution of sensor data:

1. Flattening Sensor Values: Instead of maintaining the spatial structure, all sensor readings were flattened into a single array, ensuring that information from all sensors was combined into a single dimension. This approach emphasizes the overall pattern of sensor responses rather than their spatial layout.
2. Normalization: The flattened data was normalized to the range [0, 1] to ensure uniform scaling and compatibility with CNN models.
3. Reshaping for CNN Input: The normalized data was reshaped into a format suitable for 2D processing by adding spatial dimensions and interpolated to a size of 224×224 using bilinear interpolation. This step ensures compatibility with common 2D-CNN architectures while focusing on the combined response of all sensors.

The modified approach enables the representation of sensor data as a single-channel 2D image where each pixel intensity reflects the normalized response of the sensors. This method allows the model to analyze the overall distribution of responses rather than focusing solely on predefined spatial relationships.

The changes aim to improve the representation of sensor data by emphasizing the collective sensor response instead of individual spatial relationships. This approach may better suit datasets where the spatial arrangement of sensors is less critical than the combined sensor patterns. Additionally, the single-channel format simplifies the input structure, reducing complexity without sacrificing critical information.

E. Experiment Design

The experiments were designed to evaluate the impact of Gaussian noise augmentation on classification performance under two different input representations: 1D and 2D. This comparison aimed to determine how noise-augmented data was handled in each scenario and whether the added noise contributed positively to the classification process. Augmented data was trained using two 1D-CNN architectures, adapted from [14] and [15], and multiple 2D-CNN architectures (ResNet18, ResNet34, ResNet50) with and without transfer learning. Zhou's 1D-CNN architecture featured multiple convolutional layers and max-pooling operations, optimized for extracting temporal patterns from time-series data, making it suitable for handling the multivariate nature of e-nose sensor responses. Yen's 1D-CNN, on the other hand, incorporated batch normalization layers to stabilize training with noise-augmented data and a global average pooling layer for efficient dimensionality reduction. Both models were selected for their demonstrated ability to process time-series data effectively.

For 2D-CNN architectures, the sensor data was transformed into 2D representations compatible with ResNet models. This transformation involved flattening sensor readings, normalizing them to a range of [0, 1], reshaping the data for spatial compatibility, and resizing them to a 224x224 input format using bilinear interpolation. ResNet architectures were chosen due to their ability to extract hierarchical patterns from image-like data and their robustness in handling high-dimensional inputs. Transfer learning was also applied to leverage pre-trained weights, providing a strong initialization for classification tasks.

The dataset was divided into training and testing sets, ensuring balanced representation across three tea quality classes (Q1, Q2, Q3). Gaussian noise augmentation was applied to the training set, introducing six levels of noise (0.01, 0.03, 0.05, 0.1, 0.2, and 0.3) to evaluate the effect of noise levels on model performance. Preprocessing steps included baseline manipulation using the friction method (Bardwart, 2021) to stabilize sensor responses and ensure class balance before and after augmentation.

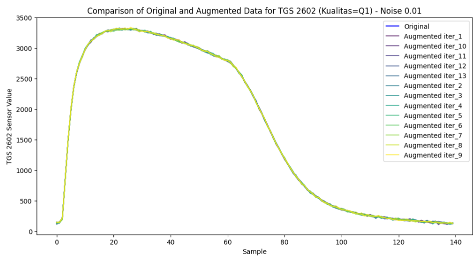
To evaluate the quality of the augmented data, statistical and visual methods were employed. MANOVA was used to test for significant differences in mean vectors between the original and augmented datasets across multiple sensor features. The Kolmogorov-Smirnov (KS) test complemented this by comparing the cumulative distributions of individual sensor responses between the original and augmented datasets, ensuring that the noise augmentation did not deviate significantly from the natural patterns of the data. Additionally, radar plots were used to visualize and compare the aroma profiles of original and augmented data across the three tea quality classes, highlighting consistency in sensor responses.

The classification performance of the models was assessed using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and confusion matrices. Accuracy measured the overall correctness of predictions, while precision and recall provided insights into the model's performance for each class. The F1-score offered a balance between precision and recall, making it particularly suitable for multiclass evaluations. Confusion matrices detailed misclassification patterns, enabling a deeper understanding of the challenges faced by the models under different input representations and noise levels.

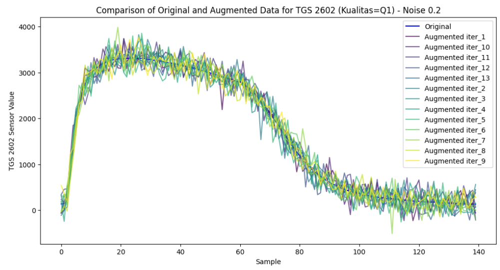
IV. RESULTS AND DISCUSSION

The results demonstrate the effectiveness of Gaussian noise augmentation in enhancing the classification performance of e-nose data, with notable differences observed between 1D and 2D input representations. While 1D-CNN models excelled in capturing temporal dependencies in the sensor data, 2D-CNN architectures, particularly those with transfer learning, showcased superior ability to leverage spatial patterns in augmented representations. Statistical evaluations using MANOVA and KS tests confirmed that the augmented data retained the core characteristics of the original aroma profiles, ensuring meaningful learning. These findings highlight the adaptability of different architectures to noise-augmented data and provide insights into optimizing e-nose-based classification systems for real-world variability.

A. Evaluation Results of Data Augmentation



(a)

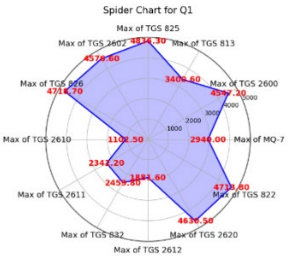


(b)

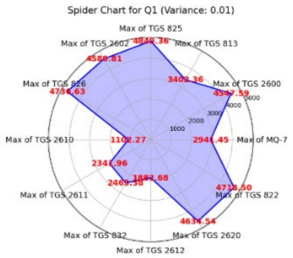
Figure 1 Augmented signal shapes: (a) Noise level 0.01; (b) Noise level 0.2

The comparison between the original and Gaussian noise-augmented sensor responses (TGS 2602) for Q1 quality under noise levels of 0.01 and 0.2 (ilustrate in Fig. 1) highlights the trade-off between preserving data integrity and introducing variability. At a noise level of 0.01, the augmented data aligns closely with the original response curve, showing minimal deviations and maintaining the overall shape and trend of the sensor readings. This indicates that low noise levels effectively add controlled variability while preserving the original data's structure, ensuring consistency in the aroma profile. In contrast, at a noise level of 0.2, the augmented data exhibits more pronounced fluctuations, particularly in regions of high sensor activity. While the general shape of the response curve remains intact, the higher noise introduces greater variability, simulating real-world disturbances but potentially reducing signal clarity.

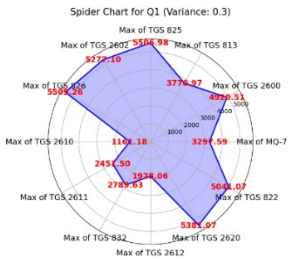
The aroma profile generated by the e-nose is utilized to identify and differentiate sample quality based on the sensitivity of sensors in detecting volatile compounds. In this context, the aroma profiles are presented as spider charts to visualize the maximum sensor readings across various gas sensors, effectively representing the aromatic components of each quality category. This visualization method facilitates a clearer and more intuitive analysis of the differences in sensor response patterns across varying levels of noise, providing insights into the extent to which noise impacts the stability of sensor readings in distinguishing the aroma profiles of each category.



(a)



(b)



(c)

Figure 2. Arome profile of Q1 quality for: (a) Original; (b) Noise level 0.01; (c) Noise level 0.3

The spider chart in Figure 2 illustrates the maximum distribution of gas sensor readings from the e-nose for Q1 quality at two levels of noise variance (0.01 and 0.3). The results demonstrate that despite the increase in noise from 0.01 to 0.3, the sensor response patterns remain stable, with minimal changes in the area of each spider chart. This indicates that the added noise does not significantly impact the aroma profile across all quality categories. Consequently, both the original and augmented data maintain consistent aroma profiles, confirming that data augmentation does not alter quality classification based on aroma profiles.

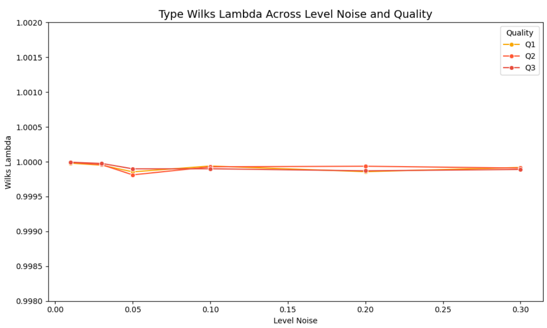
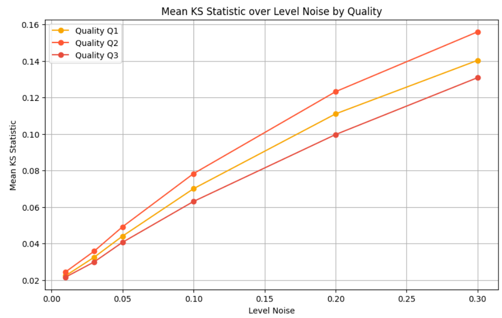
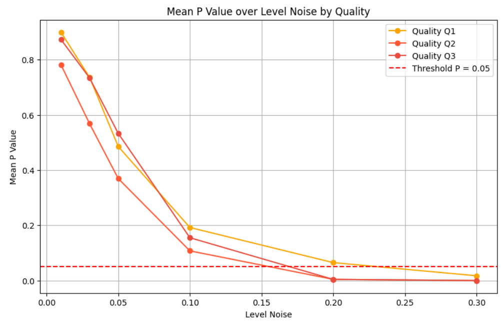


Figure 3. Wilks' Lambda values at various noise levels

The MANOVA results in Fig. 3 present Wilks' Lambda values for quality categories Q1, Q2, and Q3 across various noise levels. Consistently close to 1 with p-values > 0.05, these values indicate no significant multivariate differences between the original and augmented data distributions. In this context, Wilks' Lambda measures how well the augmented data preserves the combined variance of observed sensor variables. The near-identical distributions suggest that the Gaussian noise introduced during augmentation is subtle enough to maintain the structural integrity of the data, reflecting its ability to simulate natural variability without disrupting the overall aroma profile classification.



(a)



(b)

Figure 4. (a) Plot of the average KS statistic values across various noise levels, (b) Plot of the average P-values across various noise levels

The KS statistic increases steadily as the noise level rises from 0.01 to 0.30 across all quality categories. This indicates that higher noise levels lead to greater deviations between the cumulative distributions of the original and augmented data. The gradual increase reflects the fact that Gaussian noise, at higher levels, introduces more variability to the sensor response patterns, making the augmented data diverge further from the original distributions.

The p-values, on the other hand, exhibit a sharp decline as noise levels increase, eventually falling below the critical threshold of 0.05. This trend confirms that the augmented data becomes statistically distinguishable from the original data at higher noise levels. For lower noise levels (e.g., 0.01 and 0.03), the p-values are higher (e.g., >0.05), suggesting that the augmentation process retains a high degree of similarity to the original data. However, as the noise intensifies, the p-values drop sharply, highlighting significant differences in distribution. Gaussian noise introduces controlled randomness to the data, simulating real-world sensor variability. At lower noise levels, the added noise is subtle, primarily enriching the dataset without disrupting its statistical structure. As a result, the KS statistic remains low, and p-values stay above 0.05, signifying that the distributions are still closely aligned. At higher noise levels, however, the added variability exceeds the natural variation in the original data, leading to a breakdown in similarity in term of statistical assesment.

B. *Evaluation Results of classification using 1D-CNN*

In comparing the performance of machine learning models, understanding the role of data augmentation offers valuable insights into how models manage variability and noise in data. Data augmentation, by creating variations of the original dataset, typically enhances model robustness, accuracy, and generalization, particularly in complex real-world environments. This study evaluates the performance of two 1D CNN models, proposed by Zhou (2024) and Yen (2024), both with and without data augmentation under different noise levels, to assess the true impact of this technique.

Table. 1 Accuracy 1D-CNN, LSTM, GRU

Noise Level	Data Augmentation	Accuracy			
		1D CNN [14]	1D CNN [15]	LSTM	GRU
0.00	x	0.82	0.64	0.64	0.68
0.01	v	0.96	0.89	0.93	0.79
0.03	v	0.96	0.89	0.93	0.93
0.05	v	0.96	0.96	0.93	0.93
0.1	v	0.96	0.93	0.93	0.93
0.2	v	0.96	0.96	0.93	0.93
0.3	v	0.96	0.96	0.93	0.93

The analysis of the accuracy results for the 1D CNN models by Zhou (2024) and Yen (2024) as well as LSTM and GRU architectures, underscores the substantial influence of data augmentation on model performance (see Table 1). Without data augmentation, Zhou's 1D CNN model achieves an accuracy of 0.82 at a noise level of 0.00, while Yen's 1D CNN model, along with LSTM and GRU models, attains an accuracy of only 0.64. However, the introduction of data augmentation significantly boosts accuracy across all architectures. For the LSTM model, the use of data augmentation consistently enhances accuracy, reaching 0.93 at all noise levels from 0.01 onward, compared to 0.64 without augmentation. This indicates that LSTM's ability to capture temporal patterns is greatly supported by the noise-augmented training process. The GRU model shows similar improvements with data augmentation. While its accuracy starts at 0.79 without augmentation, it stabilizes at 0.93 across all augmented noise levels. This suggests that GRU, like LSTM, leverages augmented data effectively to maintain robust performance.

Without data augmentation, Zhou's model achieves an accuracy of 0.82 at a noise level of 0.00, while Yen's model attains an accuracy of only 0.64. However, with data augmentation, both models exhibit a marked increase in accuracy. Zhou's model consistently achieves 0.96 accuracy across all noise levels from 0.01 to 0.3, demonstrating its resilience to noise and stable performance despite varying levels of perturbation. Similarly, Yen's model benefits significantly from augmentation, with its accuracy rising to 0.96 at noise levels of 0.05 and higher, suggesting enhanced robustness and generalization.

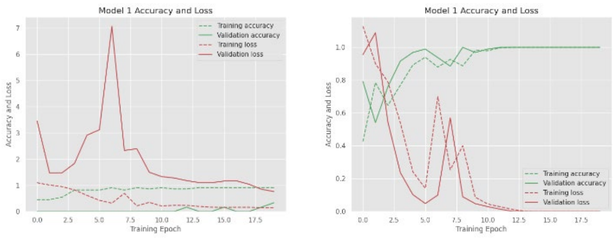
The comparative analysis between the models reveals that, in the absence of data augmentation, Zhou's model has a clear performance advantage, as evidenced by its higher accuracy. However, the application of data augmentation reduces this disparity, enabling Yen's model to achieve accuracy comparable to Zhou's. This indicates that augmentation not only improves accuracy but also serves to bridge the performance gap between models with differing baseline capabilities. Additionally, the observed fluctuations in Yen's model accuracy at lower noise levels (0.01 and 0.03, both at 0.89 accuracy) indicate an initial difficulty in coping with minor noise. Nevertheless, the model gains stability as noise levels increase, largely due to the benefits conferred by augmentation.

The experiment reveals that data augmentation plays a crucial role in enhancing the performance of Zhou's 1D CNN model (see Table 2), proven in classification report. Without augmentation (noise level 0.00), the model struggles with class Q3, as shown by lower precision, recall, and F1-score values. This suggests that class Q3 may be more challenging to classify, likely due to intrinsic characteristics of the data. When data augmentation is introduced (starting at noise level 0.01), the model demonstrates a marked improvement across all classes, especially for Q3, where the F1 score rises significantly. For noise levels 0.03 and above, the model reaches near-perfect metrics for classes Q1 and Q2, and a high level of performance for class Q3, indicating that the augmentation helps balance the model's ability to classify all three classes more effectively. The model's stability is evident as it maintains high performance across varying noise levels, with minimal degradation in metrics. Precision, recall, and F1-score mostly reach 1 or close to 1, reflecting robustness to noise. The slight fluctuation in recall for class Q3 (dipping to 0.9 at some noise levels) suggests that while augmentation has greatly improved the model's ability to generalize, there is still a slight sensitivity to certain noise levels when it comes to this particular class. Class Q1 and Q2, These classes are classified with high accuracy, precision, recall, and F1-score, consistently achieving values close to 1 across all noise levels. This suggests that these classes are easier to distinguish, and the model can generalize well even under noisy conditions. For class Q3, although there is a significant improvement with data augmentation, Q3 still shows slight vulnerabilities to noise compared to Q1 and Q2, as indicated by fluctuations in recall. This highlights that further refinements may be needed for this class to achieve consistent performance comparable to the other classes.

Table 2 Classification report using model [14]

Level Noise	Precision			Recall			F1 score		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
0.00	0.90	0.82	0.71	0.90	0.90	0.62	0.90	0.86	0.67
0.01	1	1	0.89	1	0.90	1	1	0.95	0.94
0.03	1	1	0.89	1	0.90	1	1	0.95	0.94
0.05	0.91	1	1	1	0.9	1	0.95	0.95	1
0.1	0.91	1	1	1	0.9	1	0.95	0.95	1
0.2	0.91	1	1	1	0.9	1	0.95	0.95	1
0.3	0.91	1	1	1	0.9	1	0.95	0.95	1

Based on the two provided accuracy and loss plots for different noise levels—0.00 (Fig. 4a) and 0.3 (Fig. 4b). Zhou's model contains multiple convolutional layers, which contribute to a high learning capacity. While this is beneficial for capturing complex features, it also makes the model prone to overfitting, especially when it is trained without sufficient data variability. In the noise level 0.00 scenario, the model's high capacity causes it to memorize the training set effectively, as seen by the training accuracy reaching 1.0 very quickly. However, this leads to poor generalization, evidenced by fluctuating validation accuracy and high validation loss. At noise level 0.00, the model tends to overfit, as indicated by high training accuracy and fluctuating validation accuracy and loss. This shows that the model, due to its complex architecture, is learning specific patterns in the training data that do not generalize well. At noise level 0.3, the introduction of noise acts as an effective regularizer. This results in more consistent validation accuracy and lower, stable validation loss, indicating better generalization. Zhou's architecture, with its deep layers, benefits from this augmentation because it encourages the model to learn features that are more robust to variations in the input, thus improving its performance on unseen data. Introducing noise during training at level 0.3 serves as a form of data augmentation that makes the model more robust by preventing it from overfitting to the training data.



(a) (b)

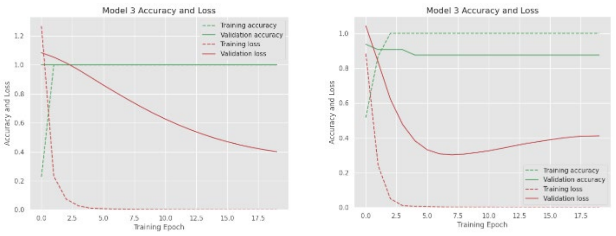
Figure 4. Accuracy and Loss of (a) No-augmentation, (b) Augmentation with noise 0.3

In the other hand, by using model Yen, 2024, for level noise 0.00, the precision for Q2 is zero, which indicates that the model is not predicting any true positive instances for this class (see Table 3). It means the model is failing to identify class Q2 in its predictions. Overall, recall is perfect for all classes, indicating that the model is sensitive and can capture most true positives, but struggles with precision—especially for Q2 and Q3—leading to imbalanced performance. Early Noise Levels (0.01, 0.03), there is a rapid improvement in precision and F1 scores for classes that were initially problematic (i.e., Q2 and Q3). The model quickly learns to correctly classify Q2 (previously missed entirely) and shows an improved F1 score for Q3. At noise level 0.05 and beyond, the model stabilizes in its performance. Precision and recall values reach consistent levels, and the F1 scores for Q1 and Q2 reach 1, indicating nearly better performance. The recall for Q3 remains slightly below 1, suggesting that Q3 remains more challenging to detect completely. However, with high precision, the model is accurate when it does classify instances as Q3, resulting in improved F1 scores. Without augmentation (noise level 0.00), Yen's model struggles particularly with Q2 and Q3. The precision for Q2 is 0, and Q3's F1 score is notably low. This indicates that the model lacks robustness when trained on non-augmented data, as it is unable to properly generalize to those classes. With noise augmentation (starting from noise level 0.01), the model shows significant improvements in precision and recall for problematic classes, especially Q2, where the precision improves from 0 to 1. This suggests that data augmentation helps the model generalize better by learning more diverse features. Yen's model benefits significantly from the introduction of data augmentation, particularly for classes Q2 and Q3, which struggled without augmentation. The addition of noise during training enhances the model's ability to correctly identify instances from these challenging classes, as seen by improvements in precision, recall, and F1 scores. The performance stabilizes at higher noise levels, indicating that augmentation helps improve generalization and robustness, particularly for complex patterns associated with Q3. Despite some remaining challenges in fully detecting Q3 instances, the model performs well overall, with augmentation effectively mitigating previous deficiencies.

Table 3 Classification report using model [15]

Level Noise	Precision			Recall			F1 score		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
0.00	0.83	0	0.5	1	0	1	0.91	1	0.67
0.01	0.83	1	0.89	1	0.7	1	0.91	0.82	0.94
0.03	1	0.82	0.88	0.9	0.9	0.88	0.95	0.86	0.88
0.05	1	1	0.89	1	0.9	1	1	0.95	0.94
0.1	1	0.9	0.89	0.9	0.9	1	0.95	0.90	0.94
0.2	1	1	0.89	1	0.9	1	1	0.95	0.94
0.3	1	1	0.89	1	0.9	1	1	0.95	0.94

Based on the accuracy and loss plots for Model 3 at different noise levels—0.00 (Fig. 5a) and 0.3 (Fig. 5b). At noise level 0.00, the model overfits the training data, as indicated by the training accuracy reaching 1.0 almost immediately, while validation accuracy lags behind. The introduction of noise level 0.3 reduces this overfitting by preventing the model from memorizing the training data and forcing it to learn more robust features. Batch Normalization layers in Yen's architecture are instrumental in stabilizing the training process, especially in the presence of noise. By normalizing activations, these layers allow the model to learn efficiently without being significantly affected by the variability introduced by noise. With noise level 0.3, Batch Normalization helps the model converge steadily without overfitting, contributing to the improved validation accuracy and reduced gap between training and validation performance. The use of GlobalAveragePooling1D before the fully connected layers contributes to reducing overfitting by summarizing spatial features. This is particularly effective when noise is added, as it helps maintain focus on the most significant features rather than all the learned details. This pooling strategy aligns well with the trend observed at noise level 0.3, where the model performs better on unseen data, showing more consistent and reliable generalization. Despite the slight increase in validation loss, the validation accuracy remains stable, suggesting that while the model is overfitting to some extent, it still maintains reasonable performance on validation data. The complex architecture of Yen's architecture with multiple convolutional layers, each having 128 filters, provides a high capacity for learning, which may cause overfitting, especially if the training set is not sufficiently diverse or large. The lack of regularization techniques such as dropout could also contribute to overfitting. Although Batch Normalization helps stabilize training and potentially reduces overfitting to some degree, adding dropout layers could further enhance generalization.



(a) (b)

Figure 5. Accuracy and Loss of (a) No-augmentation, (b)Augmentation with level noise 0.3

C. Evaluation Results of classification using 2D-CNN

To assess the impact of Gaussian noise augmentation on e-nose data classification, the performance of three 2D CNN architectures—ResNet 18, ResNet 34, and ResNet 50—was evaluated across varying noise levels. These models were tested both with and without transfer learning to examine their ability to generalize under noisy conditions. The results, summarized in Table 4, reveal notable differences in performance across architectures and noise levels, highlighting the role of transfer learning and the robustness of each model. The following detailed analysis explores the intricate relationships between noise levels, model architectures, and the influence of transfer learning on classification performance.

Table 4 Accuracy using 2D CNN

Noise Level	Transfer Learning	Accuracy		
		ResNet 18	ResNet 34	ResNet 50
0.00	True	30.09	36.80	32.57
0.00	False	34.71	28.80	33.52
0.01	True	92.66	98.66	75.23
0.01	False	89.71	92.28	56.19
0.03	True	87.76	93.76	75.85
0.03	False	91.80	89.38	66.28
0.05	True	91.09	94.09	70.95
0.05	False	94.95	90.80	54.14
0.1	True	91.28	92.42	86.95
0.1	False	93.85	78.80	51.80
0.2	True	85.71	94.76	64.33
0.2	False	86.33	79.42	47.09
0.3	True	76.61	89.80	55.42
0.3	False	79.14	77.47	49.52

ResNet architectures (ResNet 18, 34, and 50) leverage pre-trained weights from large-scale datasets like ImageNet, enabling efficient transfer learning by providing robust, generalizable feature representations. ResNet 18 and ResNet 34, with their shallower structures, demonstrate greater robustness to noise and perform consistently better than ResNet 50 across most noise levels. For instance, at a noise level of 0.2, ResNet 18 achieves an accuracy of 85.71% with transfer learning, while ResNet 50 achieves only 64.33%. Similarly, at a noise level of 0.05, ResNet 34 achieves 94.09% accuracy with transfer learning, significantly outperforming ResNet 50, which achieves only 70.95% under the same conditions. This pattern highlights that simpler architectures, like ResNet 18 and ResNet 34, are better suited for handling noisy data due to their reduced complexity and lower susceptibility to overfitting. In contrast, ResNet 50, with its deeper architecture and higher parameter count, struggles to generalize effectively, especially at higher noise levels. For example, at a noise level of 0.3, ResNet 50 achieves only 55.42% accuracy with transfer learning and drops further to 49.52% without transfer learning. This suggests that while deeper architectures like ResNet 50 offer greater capacity, they require additional regularization or larger datasets to handle noisy environments effectively.

Transfer learning emerges as a crucial factor in improving the performance of all models across noise levels. By leveraging pre-trained weights, transfer learning enables the models to generalize better in the presence of noise. For instance, at a noise level of 0.01, ResNet 18 improves from 89.71% to 92.66% with transfer learning, while ResNet 34 improves from 92.28% to 98.66%, and ResNet 50 improves significantly from 56.19% to 75.23%. This demonstrates the ability of transfer learning to extract robust features that adapt well to noisy data. Without transfer learning, all models exhibit significantly lower accuracy, especially at higher noise levels. For instance, at a noise level of 0.2, ResNet 50 achieves only 47.09% accuracy without transfer learning, compared to 64.33% with transfer learning. These results underline the importance of transfer learning as a key strategy for improving generalization and mitigating the effects of noise in 2D CNN models.

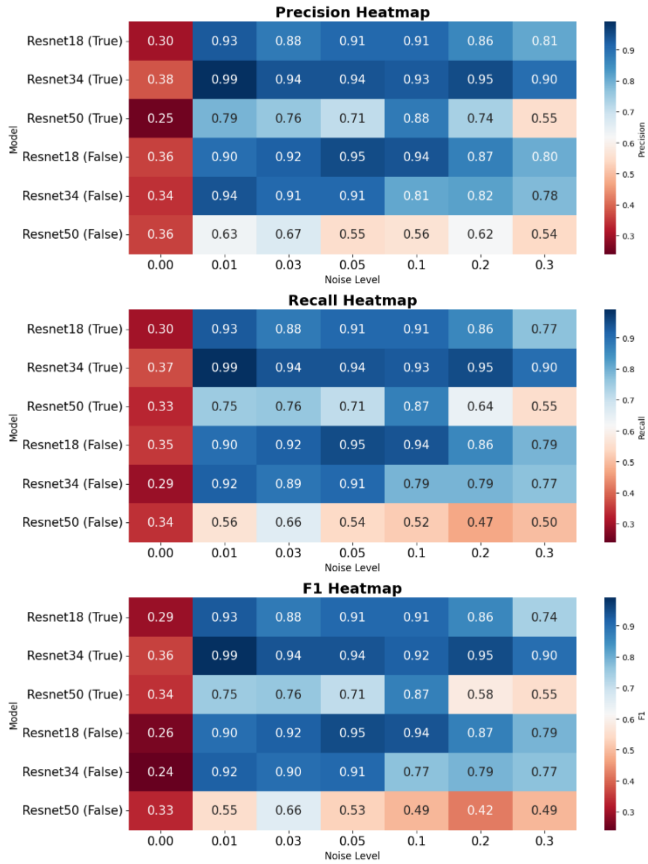


Figure 6 Heatmap of precision, recall, and F1-score for all model each level noise

The performance of different ResNet architectures (ResNet18, ResNet34, and ResNet50) with and without transfer learning was evaluated across varying noise levels using three key metrics: precision, recall, and F1-score (see Fig. 6). The reported values are macro averages, representing the arithmetic mean across three classes (Q1, Q2, Q3), which ensures that all classes are weighted equally, regardless of their support.

The results reveal that ResNet34 with transfer learning consistently outperforms other models across all metrics and noise levels. This superiority is most evident at lower noise levels (e.g., 0.01), where ResNet34 achieves near-perfect scores (precision: 0.99, recall: 0.99, F1: 0.99). ResNet18 with transfer learning also performs robustly, particularly at higher noise levels, where it retains relatively high metric values. In contrast, ResNet50 without transfer learning exhibits the poorest performance, with significant drops in all metrics, especially at higher noise levels (e.g., 0.3). This underscores the challenge of handling noise in complex models like ResNet50.

The role of transfer learning is evident in mitigating performance degradation across all models. Models with transfer learning demonstrate greater resilience to noise, maintaining higher scores compared to their non-transfer counterparts. For example, at a noise level of 0.3, ResNet18 with transfer learning achieves a precision of 0.81, whereas the same model without transfer learning scores only 0.80. Similarly, ResNet50 with transfer learning shows smaller declines compared to its non-transfer version, highlighting the importance of leveraging pre-trained weights for better generalization.

Noise levels play a crucial role in model performance. As noise increases, all models experience a decline in precision, recall, and F1-score, reflecting the growing difficulty of differentiating between true signals and noise. The decline is particularly noticeable in recall, which measures the model's ability to identify all true positives. For example, ResNet50 without transfer learning sees its recall drop from 0.56 at a noise level of 0.01 to 0.47 at 0.3, highlighting the challenge of maintaining sensitivity in noisy environments.

At low noise levels (e.g., 0.01), all models perform well, especially those using transfer learning, because the data is clean and easier to analyze, making it simpler for the models to identify patterns and classify correctly. However, as noise levels get higher (e.g., 0.2 or 0.3), the models struggle to handle the noise, and their performance drops. This effect is worse for more complex models like ResNet50, which are more likely to overfit to the noisy data. In contrast, simpler models like ResNet18 and ResNet34 handle noise better because they are less complex and less likely to overfit, making them more reliable in noisy situations.

Recall shows the largest drop across models, particularly at higher noise levels. This is because recall measures the ability to identify all true positives. As noise increases, the models struggle to differentiate true signals from noise, leading to an increase in false negatives. Precision, on the other hand, remains relatively stable as it depends on the proportion of true positives among the predicted positives, which is less affected by the rise in false negatives. Consequently, the F1-score, as a harmonic mean of precision and recall, reflects these declines but remains more balanced between the two.

ResNet34 with transfer learning emerges as the most effective architecture due to its balance between complexity and feature extraction capacity, enabling it to perform well across noise levels. Noise significantly impacts model performance, with higher noise levels exacerbating misclassification and lowering metrics, especially for more complex models like ResNet50. Leveraging transfer learning proves to be a critical factor in improving generalization and mitigating the effects of noise, particularly in challenging scenarios.

V. CONCLUSION AND FUTURE WORK

The addition of Gaussian noise as data augmentation offers a promising non-disruptive approach for food analysts to enhance black tea classification using e-nose. The addition of Gaussian noise through the formula introduces controlled variability that simulates real-world conditions while preserving the core characteristics of the original data. Gaussian noise augmentation proved to be a valuable tool in enhancing the robustness of e-nose data classification, as confirmed by statistical evaluation using MANOVA and KS tests. The statistical results demonstrated that augmented data maintained the original data's structural integrity at lower noise levels, ensuring realistic variability while preserving core characteristics of aroma profiles. However, at higher noise levels, the KS statistic and declining p-values indicated increased deviations, suggesting a trade-off between variability and statistical similarity. These findings emphasize that carefully calibrated noise levels are essential for achieving a balance between data augmentation benefits and the preservation of data integrity.

The evaluation of 1D-CNN models highlighted the significant role of noise-augmented data in improving model performance across varying noise levels. Without augmentation, models struggled with lower accuracy and inconsistent classification, particularly for challenging classes such as Q3. Data augmentation resulted in consistent improvements in precision, recall, and F1 scores, enabling better generalization and balanced performance across all classes. The performance stability across noise levels showcased the adaptability of 1D-CNN architectures, with augmentation helping mitigate overfitting and enhancing robustness, particularly for models with deeper architectures like Zhou's 1D-CNN.

In the 2D-CNN classification, transfer learning emerged as a critical factor for handling noise-augmented data effectively. ResNet34 with transfer learning demonstrated superior performance, balancing complexity and generalization capacity across noise levels. Simpler architectures like ResNet18 were also robust under noisy conditions, whereas the deeper ResNet50 faced challenges, particularly at higher noise levels, due to overfitting. The results underscore the importance of leveraging pre-trained weights to enhance noise resilience, with transfer learning significantly improving precision, recall, and F1 scores. Collectively, these findings demonstrate that noise augmentation and appropriate architectural strategies can optimize e-nose classification systems for real-world variability.

Future work could explore several directions to build upon these findings. First, investigating advanced data augmentation techniques beyond Gaussian noise, such as adversarial noise or sensor-specific augmentation, could provide deeper insights into enhancing model robustness under diverse real-world conditions. Second, optimizing deeper architectures like ResNet50 by incorporating regularization techniques, such as dropout or weight decay, or exploring lightweight modifications to existing architectures, could help mitigate overfitting while maintaining high performance in noisy environments. Finally, expanding the scope to include evaluations across additional noise levels or real-world datasets could further validate the effectiveness of noise augmentation and architectural strategies in practical applications.

ACKNOWLEDGMENT

We would like to extend our heartfelt gratitude to Mr. Danang Lelono from Electronic and Instrumention of Gadjah Mada University for generously providing the dataset used in this research. His invaluable contributions have been instrumental in enabling the exploration and evaluation of e-nose technology for black tea classification. Additionally, we are deeply thankful for the computational resources provided by the Tambora server at the School of Electrical Engineering and Informatics of Institute of Technology Bandung, which significantly supported the processing and analysis of data in this study.

REFERENCES

[1] Z. B. Zhang *et al.*, "Understanding the Origin and Evolution of Tea (Camellia sinensis [L.]): Genomic Advances in Tea," Apr. 01, 2023, *Springer*. doi: 10.1007/s00239-023-10099-z.

[2] M. Li, T. Pan, and Q. Chen, "Estimation of tea quality grade using statistical identification of key variables," *Food Control*, vol. 119, Jan. 2021, doi: 10.1016/j.foodcont.2020.107485.

[3] S. S. Q. Rodrigues, L. G. Dias, and A. Teixeira, "Emerging Methods for the Evaluation of Sensory Quality of Food: Technology at Service," *Current Food Science and Technology Reports*, vol. 2, no. 1, pp. 77–90, Jan. 2024, doi: 10.1007/s43555-024-00019-7.

[4] A. Bhargava, A. Bansal, V. Goyal, and P. Bansal, "A review on tea quality and safety using emerging parameters," Apr. 01, 2022, *Springer*. doi: 10.1007/s11694-021-01232-x.

[5] K. O. Kombo *et al.*, "Enhancing classification rate of electronic nose system and piecewise feature extraction method to classify black tea with superior quality," *Sci Afr*, vol. 24, p. e02153, 2024, doi: https://doi.org/10.1016/j.sciaf.2024.e02153.

[6] H. Anwar, T. Anwar, and S. Murtaza, "Review on food quality assessment using machine learning and electronic nose system," *Biosens Bioelectron X*, vol. 14, p. 100365, 2023, doi: https://doi.org/10.1016/j.biosx.2023.100365.

[7] D. Lelono, L. S. Adi, A. Dharmawan, J. E. Istiyanto, and M. I. Ananta Timur, "Classification of the Coffee Roasting Level Based on Electronic Nose," in *Proceedings - 2022 8th International Conference on Science and Technology, ICST 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICST56971.2022.10136263.

[8] C. Fang *et al.*, "Smart Electronic Nose Enabled by an All-Feature Olfactory Algorithm," *Advanced Intelligent Systems*, vol. 4, no. 7, Jul. 2022, doi: 10.1002/aisy.202200074.

[9] Y. Jiang, B. Drescher, and G. Yuan, "A GAN-Based Multi-Sensor Data Augmentation Technique for CNC Machine Tool Wear Prediction," *IEEE Access*, vol. 11, pp. 95782–95795, 2023, doi: 10.1109/ACCESS.2023.3311269.

[10] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," Jul. 01, 2021, *Public Library of Science*. doi: 10.1371/journal.pone.0254841.

[11] M. Arslan, M. Demirci, M. Guzel, and S. Ozdemir, "SMOTE and Gaussian Noise Based Sensor Data Augmentation."

[12] B. Choi, H. Kim, and S. H. Jeon, "LSTM-Based Time Series Forecasting of Pulmonary Function Test for COPD Early Diagnosis," *Journal of Korean Institute of Communications and Information Sciences*, vol. 49, no. 3, pp. 346–355, Mar. 2024, doi: 10.7840/kics.2024.49.3.346.

[13] K. Brudzewski, S. Osowski, K. Wolinska, and J. Ulaczyk, "Smell similarity on the basis of gas sensor array measurements," *Sens Actuators B Chem*, vol. 129, no. 2, pp. 643–651, Feb. 2008, doi: 10.1016/j.snb.2007.09.050.

[14] Y. Zhou *et al.*, "Enhanced Gas Recognition of Electronic Nose Using 1D Convolutional Neural Network with Savitzky-Golay Filter," *IEEE Sens J*, Apr. 2024, doi: 10.1109/JSEN.2024.3363698.

[15] Z. Ye, Y. Li, R. Jin, and Q. Li, "Toward Accurate Odor Identification and Effective Feature Learning with an AI-Empowered Electronic Nose," *IEEE Internet Things J*, vol. 11, no. 3, pp. 4735–4746, Feb. 2024, doi: 10.1109/JIOT.2023.3299555.

[16] J. Bjorck, C. Gomes, B. Selman, and K. Q. Weinberger, "Understanding Batch Normalization."

[17] H. Mahdavi, R. Goldoust, and S. Rahbarpour, "A 2D-CNN Architecture for Improving the Classification Accuracy of an Electronic Nose with Different Sensor Positions," in *2022 12th International Conference on Computer and Knowledge Engineering, ICCKE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 452–457. doi: 10.1109/ICCKE57176.2022.9960107.

[18] Y. K. Oh, C. Lim, J. Lee, S. Kim, and S. Kim, "Multichannel convolution neural network for gas mixture classification," *Ann Oper Res*, 2022, doi: 10.1007/s10479-022-04715-2.

[19] G. Wei, G. Li, J. Zhao, and A. He, "Development of a LeNet-5 gas identification CNN structure for electronic noses," *Sensors (Switzerland)*, vol. 19, no. 1, Jan. 2019, doi: 10.3390/s19010217.

[20] D. Lelono, K. Triyana, S. Hartati, and J. E. Istiyanto, "Development of Electronic Nose with Highly Stable Sample Heater to Classify Quality Levels of Local Black Tea," vol. 7, no. 2, 2017.