International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Instrumentation Mesure Métrologie
Vol. 23, No. 4, August, 2024, pp. 297-306
Journal homepage: http://iieta.org/journals/i2m

A Comparative Analysis of Early Fusion Architectures for Multimodal Gas Detection Using Machine Learning Models

Greeshma Arya[1*], Ashish Bagwari[2], Sanskriti Agarwal[1], Jasleen Kaur Aneja[1], Ciro Rodriguez[3]

[1] Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi 110006, India
[2] Department of Electronics and Communication Engineering, Uttarakhand Technical University, Dehradun 248007, India
[3] Department of Software Engineering, Universidad Nacional Mayor de San Marcos (UNMSM), Lima 15081, Peru

Corresponding Author Email: crodriguezro@unmsm.edu.pe; greeshmaarya@igdtuw.ac.in

https://doi.org/10.18280/i2m.230405

Received: 9 April 2024
Revised: 11 June 2024
Accepted: 1 August 2024
Available online: 23 August 2024

Keywords:
multi-sensor data fusion, gas classification, machine learning, early fusion, VGG19, Random Forest Classifier

ABSTRACT

Single-sensor gas detection models often lack robustness and accuracy, hindering safety and security. To enhance the accurate classification performance data from seven sensors along with thermal camera images has been used in this study, to train the model. The dataset focuses on four classes: Smoke, Perfume, No Gas and Mixture of Smoke and Perfume. Data from various sources capture different perspectives that enhance classification of the trained model, hence, early fusion technique was adopted to combine the extracted features, for an improved feature space. The sensor data undergoes preprocessing to normalize and remove noise. VGG16 model was used to extract image features. The fused data then acted as an input for the machine learning models for classification Among the tested models (SVM, Random Forest Classifier, and KNN), the Random Forest model achieved the best validation accuracy of 96.41%, outperforming SVM (94.22%) and KNN (94.53%). This approach demonstrates the effectiveness of multi-sensor data fusion for enhanced gas detection with high accuracy, potentially improving response times and reducing false alarms.

1. INTRODUCTION

Gas and smoke detection plays a critical role in ensuring safety in residential, commercial, and industrial settings. Early detection of these threats is crucial for enabling timely evacuation and emergency response. According to data from the Accidental Deaths and Suicides in India (ADSI), an average of 35 fire-related deaths occurred daily between 2016 and 2020 [1]. Carbon monoxide poisoning is also a concern in India. While specific data on CO poisoning incidents is limited, a study [2] highlights the growing problem of indoor air pollution, which can be a major source of CO exposure. Given the dangers posed by gas leaks, relying on human intervention alone is not feasible. Instead, there is a critical need for machines to swiftly and accurately detect and respond to gas leaks to ensure safety. Therefore, timely identification of gas leaks is paramount for preventing potential disasters.

An effective strategy to achieve this is by integrating machine learning methodologies into gas detection and classification systems. By employing machine learning algorithms, these systems can analyze large amounts of sensor data to swiftly detect and categorize gas leaks or smoke presence based on their unique signatures. This approach not only enhances the efficiency of detection but also enables proactive measures to be taken promptly, minimizing the risk of catastrophic events.

Traditional gas detection methods primarily rely on metal oxide semiconductor (MOS) sensors. While these sensors offer advantages like affordability and ease of use, they suffer from limitations. Their sensitivity to environmental factors like temperature, humidity, and even minor fluctuations in surrounding air composition can lead to false alarms or reduced accuracy in gas identification. Therefore, relying solely on data collected from these sensors could bring discrepancies in the observations.

This research explores the potential of multimodal data fusion as an effective approach for gas classification. Multimodal data fusion leverages information from multiple sources to create a richer and more informative dataset, potentially overcoming the limitations inherent in single-sensor methods. In this work, we experiment and make observations on a multimodal dataset named Multimodal Gas Data. This dataset goes beyond traditional sensor data by incorporating thermal images captured alongside measurements from seven distinct MOS gas sensors. By combining these two data modalities, we aim to extract a more comprehensive signature of the target gas, leading to improved classification accuracy.

To unlock the potential of the multimodal dataset, we employ VGG16, a well-established convolutional neural network architecture, for feature extraction from the thermal images and sensor data. We then implement an early fusion approach, where the extracted features from the thermal images are combined with the raw sensor data to create a unified feature vector. Finally, we evaluate the performance of three machine learning models - Random Forest Classifier, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) - in classifying four distinct gas classes: "Smoke", "Perfume", "No Gas" and a "Mixture of Perfume and Smoke." By comparing the performance of these models, we aim to identify the most effective approach for gas classification using the Multimodal Gas Data dataset. Through this comprehensive investigation, we hope to demonstrate the efficacy of multimodal data fusion in enhancing gas classification accuracy, paving the way for the development of more robust and reliable gas detection systems.

2. LITERATURE REVIEW

Gas classification has garnered significant attention due to its wide range of applications including environmental monitoring, industrial safety and healthcare, as undetected gas leaks contribute to air and water pollution, potentially causing respiratory illnesses and ecological damage, while inaccurate detection could lead to misdiagnosis in healthcare and safety risks for workers in industrial settings. Various studies have explored different methodologies and techniques to enhance the accuracy and efficiency of gas classification systems. However, the efficacy of gas classification systems relying solely on sensor data is often hindered by factors such as sensor cross-sensitivity and environmental variability. As such, recent research efforts have delved into incorporating traditional sensor-based approaches with emerging technologies like thermal imaging to increase classification accuracy. Furthermore, the fusion of data from multiple modalities has garnered significant interest, drawing from the rich landscape of multimodal data fusion techniques employed across various domains such as computer vision and biometrics.

To address these limitations, researchers are increasingly turning to machine learning for gas classification. Use of machine learning models provides better accuracy and ability to handle complex sensor data, in case of gas detection. Pardo and Sberveglieri [3] proposed a gas classification system using an array of MOS sensors and a Support Vector Machine (SVM) for pattern recognition. Their approach achieved high accuracy in differentiating various volatile organic compounds (VOCs). Peng et al. [4] propose GasNet, a novel Deep CNN architecture for gas classification. This work demonstrates that a DCNN approach can achieve superior accuracy compared to traditional machine learning methods like SVM and MLP for electronic nose data classification. Oh et al. [5] explore using machine learning with a four-sensor metal oxide gas sensor array for CO and ethanol detection. Their work highlights the effectiveness of combining unsupervised learning (K-Means clustering) with supervised learning (DNNs, CNNs) to achieve high accuracy gas classification, even with a limited number of sensors.

Thermal imaging has emerged as a promising technology for gas classification as it can capture spatial distribution patterns of gases based on their thermal properties. By analyzing temperature variations in the infrared spectrum [6], thermal imaging can complement traditional sensor-based approaches by providing additional spatial information for gas localization and characterization. Xiong et al. [7] combines thermal imaging and deep learning to detect underground natural gas micro-leakage by analyzing the temperature changes in stressed vegetation above the leak. They achieve high accuracy in identifying these stressed areas using a convolutional neural network model.

The utilization of only gas sensors may encounter challenges stemming from their susceptibility to cross-sensitivity, environmental interferences, and limited selectivity, which can compromise the discriminatory power required for precise gas identification. Similarly, sole reliance on thermal imaging for gas detection confronts obstacles such as difficulties in effectively discriminating between gases with similar thermal signatures and potential signal distortions induced by ambient temperature variations. This highlights the necessity for analyzing data from multiple sources simultaneously.

In multimodal data fusion, information from multiple sources or modalities is integrated to enhance the overall understanding or performance of a system. This concept has found applications in various domains including computer vision, biometrics, and sensor networks. A study by Chen et al. [8] explores neural biomarkers for bipolar disorder and schizophrenia through multimodal neuroimaging analysis. Results underscore the significance of multimodal fusion in distinguishing between the two disorders. By combining complementary information from different modalities, multimodal data fusion can improve robustness, reliability, and accuracy of classification or detection tasks. Another research conducted by Dyrba et al. [9] investigates the utility of multimodal MRI data in automated image diagnostics for Alzheimer's disease (AD). While single modalities show promising classification performance, multimodal fusion helps comprehensively assess AD-related brain alterations, contributing valuable insights for future diagnostic approaches.

For thermal imaging, an important step is executing feature extraction. The study [10] investigates target recognition in thermal infrared images. Through deep CNN-based feature extraction using pre-trained models like AlexNet and VGG19, it demonstrates enhanced performance, with VGG19_fc6 architecture achieving a notable 6% accuracy improvement over existing methods on FLIR thermal infrared datasets.

Several studies have explored the use of multimodal data fusion techniques for gas classification tasks. Through a gas detection system case study, Rahate et al. [11] demonstrate the effectiveness of multimodal co-learning for robustness, showing superior performance compared to traditional fusion methods. Narkhede et al. [12] proposed a deep learning approach that fuses data from gas sensor arrays and thermal cameras for gas classification. They achieved higher accuracy compared to using individual sensors alone. Another research [13] employed intermediate and multitask fusion to achieve similar results.

By combining sensor data with additional modalities such as thermal imaging, acoustic signals, or spectroscopic measurements, researchers aim to overcome the limitations of sole sensor-based data and improve the overall performance of gas classification systems. Techniques such as feature-level fusion, decision-level fusion, and deep learning-based fusion have been investigated to effectively integrate information from multiple sources and achieve enhanced gas discrimination capabilities.

3. DATA COLLECTION

This research paper utilizes the Multimodal Gas Dataset as the foundation for conducting experiments and generating analytical insights. The dataset serves as the cornerstone of our study, enabling us to investigate various aspects related to gas detection. It consists of two modalities, a gas sensor dataset and a thermal image dataset. The gas sensor dataset comprises of three main classes of gases: “Perfume”, “Smoke”, and “Mixture” and “No Gas”. Seven metal oxide semiconductor (MOS) gas sensors, MQ2, MQ3, MQ5, MQ6, MQ7, MQ8, MQ135, were used by the dataset. The image dataset comprises of 6400 images captured using a thermal imaging camera.

The data collection process done by Narkhede et al. [12] involves, the sensor array and thermal camera subjected to two primary sources for collection of data, primary being burnt incense sticks and fragrance from a deodorant spray. It's important to acknowledge that this choice of gas sources limits the dataset's scope in terms of gas type variety. Hence, future work could work more towards expanding the dataset to encompass other commonly encountered gas types.

While the described work establishes a foundation for multimodal gas detection using the chosen sensor configuration, further exploration is necessary to enhance the generalizability of the findings. The current dataset focuses on capturing variations in gas concentration by implementing different release intervals (15 seconds, 30 seconds, and 45 seconds) during data collection [12]. However, incorporating a broader spectrum of gas concentrations within the dataset could improve the model's robustness in real-world scenarios with varying gas levels.

Another aspect to consider for future work is the influence of environmental conditions on sensor readings and thermal signatures. The referenced work mentions maintaining a "neutral environment" for the "No Gas" class, but doesn't elaborate on the control of environmental factors (temperature, humidity) during data collection [12]. Investigating the impact of these factors could provide valuable insights into the real-world applicability of the proposed approach. By acknowledging these limitations and outlining potential areas for future exploration, we aim to contribute to the development of more versatile and adaptable multimodal gas detection systems.

Before metal oxide sensors became prevalent, various gas sensing techniques were utilized. These included visual methods using chemical reagents, catalytic combustion sensors, electrochemical sensors for toxic gases, thermal conductivity sensors for industrial leak detection, and infrared absorption spectroscopy for selective gas detection. While each had its advantages, metal oxide sensors emerged as superior, the reason being their high responsiveness, low cost, high speed and versatility in gas sensing.

Metal oxide gas sensors operate based on the principle of conductivity changes in the presence of target gases. They consist of a semiconductor material typically composed of metal oxides like tin dioxide (SnO_2), zinc oxide (ZnO), or tungsten oxide (WO_3). When the sensor is exposed to a gas, molecules from the gas are adsorbed onto the surface of the metal oxide, altering the conductivity of the material. This change in conductivity is due to the interaction between the gas molecules and the surface of the metal oxide, which modifies the number of charge carriers in the material. As a result, the electrical resistance of the sensor changes, which can be measured and correlated with the concentration of the target gas.

Thermal imaging cameras operate on the principles of infrared radiation detection, a fundamental aspect of electromagnetic wave propagation. These cameras, unlike their visible light counterparts, exploit the inherent thermal emissions of objects, rendering them capable of functioning in environments devoid of external illumination, such as low-light conditions or absolute darkness. The temperature of an object influences the intensity of the infrared radiation emitted by it. The core of thermal imaging cameras is a detector array that produces electrical signals from captured infrared radiation, which are then transformed to create a thermal image, where different colors or shades of gray correspond to varying temperatures. Hotter objects appear brighter in the image, while cooler objects appear darker.

In the context of gas classification, thermal cameras can provide valuable information that might not be visible using a standard camera. Since different gases may have slightly different thermal properties, the thermal image can potentially help distinguish between the presence or absence of gas, or even between different gas types. This information, combined with data from the metal oxide gas sensors, can contribute to more robust gas classification.

It is important to acknowledge that both gas sensor readings and thermal image features have inherent limitations. Gas sensors can exhibit cross-sensitivity to certain gases, and their readings can be influenced by environmental factors like temperature and humidity. To mitigate these effects, we employed regular calibration procedures and data pre-processing techniques that accounted for potential environmental variations. Thermal image features can also be affected by background objects and variations in material emissivity. To address this, background subtraction techniques were implemented during image pre-processing. Additionally, the resolution of the thermal camera limits the ability to capture very small or diffuse gas leaks. Future studies could explore incorporating additional sensor modalities or higher resolution thermal cameras to potentially address these limitations.

4. METHODOLOGY

This section describes the methodological approach employed for processing sensor data, thermal images, and applying early fusion with ML models for gas classification.

4.1 Data preprocessing

The methodology encompasses data preprocessing for both sensor data and thermal images, followed by feature extraction from the images.

4.1.1 Sensor data preprocessing

The following subsection details how the sensor data was preprocessed. The dataset had no missing or unmatching values, hence, no missing values were handled. To mitigate any bias introduced by vastly different scales by the algorithms to be used, normalization and scaling were done using standard scaling technique. This technique utilizes the mean and standard variation, by subtracting the mean and dividing it by the standard variance, resulting with zero mean and unit variance. Standard scaling was chosen over other techniques like min-max scaling because it aligns better with the assumptions of many machine learning algorithms, particularly those based on Euclidean distance (e.g., k-Nearest Neighbors, Support Vector Machines).

The data was initially reshaped into a two-dimensional array with a single column for efficient normalization on individual sensors using the chosen technique. Subsequently, the data was flattened back into a one-dimensional array for compatibility with downstream processing steps, as shown in Figure 1.

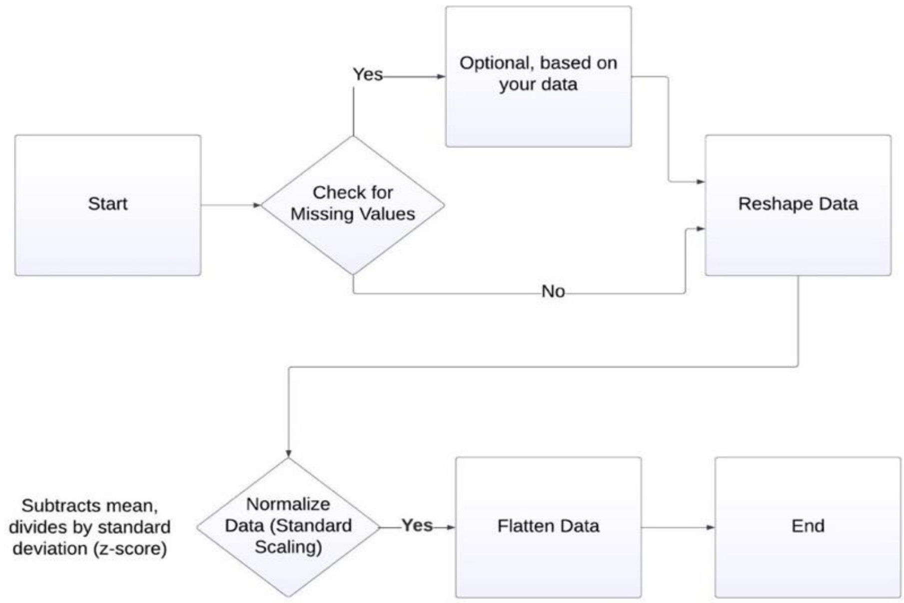


Figure 1. Sensor data preprocessing flow

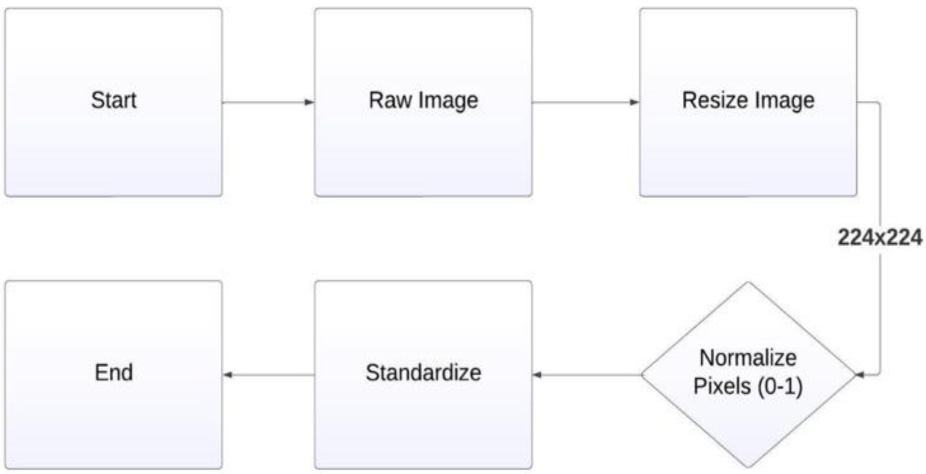


Figure 2. Image data preprocessing flow

4.1.2 Image data preprocessing

Images captured from thermal camera are preprocessed in this step to prepare the data for feature extraction, as shown in Figure 2. The image is resized to 224×224 using OpenCV resize method. This dimension is used because it complies with pre-trained model to be used in feature extraction. The image pixels are normalized between 0 and 1 by dividing using 255, to standardize the data for better training convergence. The study could have leveraged CNN for thermal image denoising, but it requires significant computation resources. This study focuses on simpler normalization techniques for preprocessing due to computational efficiency.

4.2 Feature extraction

The pre-trained CNN model, VGG16 is used for image classification. VGG16 utilizes the large ImageNet pre-trained weights. These pre-trained weights encode valuable features learned from a massive collection of labeled images. These weights can effectively extract significant features from thermal images, even though ImageNet is not trained for gas detection.

The VGG16 architecture uses 3×3 convolutional filters with ReLU (Rectified Linear Unit) activations. The first layer involves the preprocessed thermal image with 224×224 dimensions. VGG16 contains five convolutional blocks, each block having two or three convolutional layers followed by a Max Pooling Layer [14]. These layers extract feature maps from the input. The convolutional layers learn increasingly intricate features through their stacked configuration, allowing the network to capture hierarchical patterns within the image data [15]. These layers down sample the feature maps by taking the maximum value within a specific window (often 2×2). This reduces the spatial resolution of the data while retaining important features and controlling model complexity. The final layer, that is after the processing of convolutional layer and before fully connected layer, gives the output of final extracted features [14]. These feature maps typically have a lower spatial resolution compared to the original image due to the pooling operations. However, they contain a compressed representation that captures the most significant information from the image. Figure 3 shows the early fusion architecture for feature extraction, this architecture draws inspiration from the deep learning framework proposed by Asskali [16] for elbow joint effusion classification. However, the final fully connected layer is omitted to facilitate the extraction of features from the thermal images.

4.3 Data fusion

After the preprocessing and feature extraction of sensor data and thermal images, the data is fused using early fusion technique. In late fusion, as illustrated in Figure 4(a), each modal is processed separately and then fed into a different classifier suitable to the given modal. The results of both classifiers are then combined. The early fusion technique involves fusing the features of multimodal data post feature extraction into one feature vector before feeding into single decision-making model as shown in Figure 4(b).

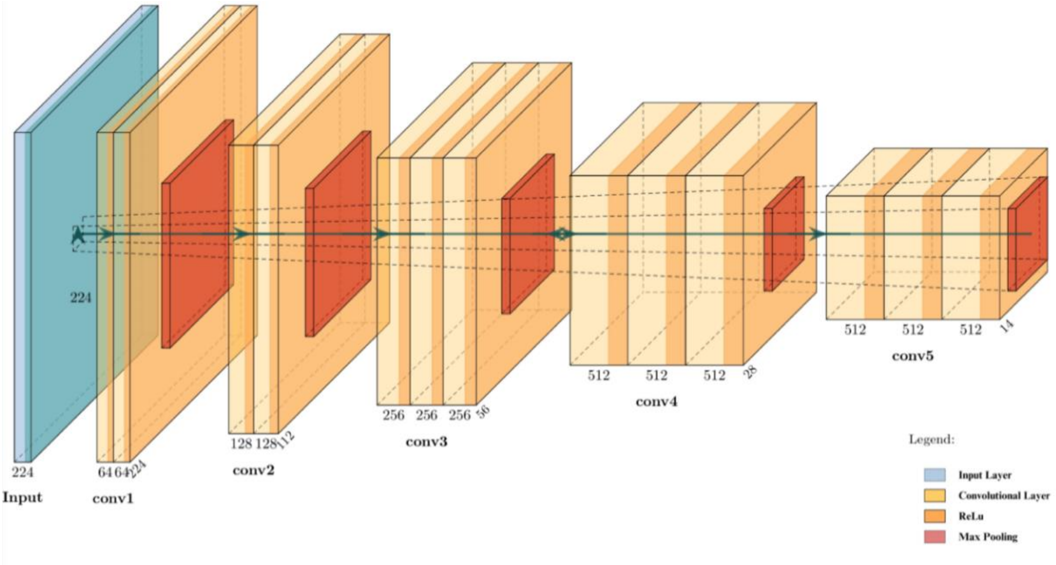
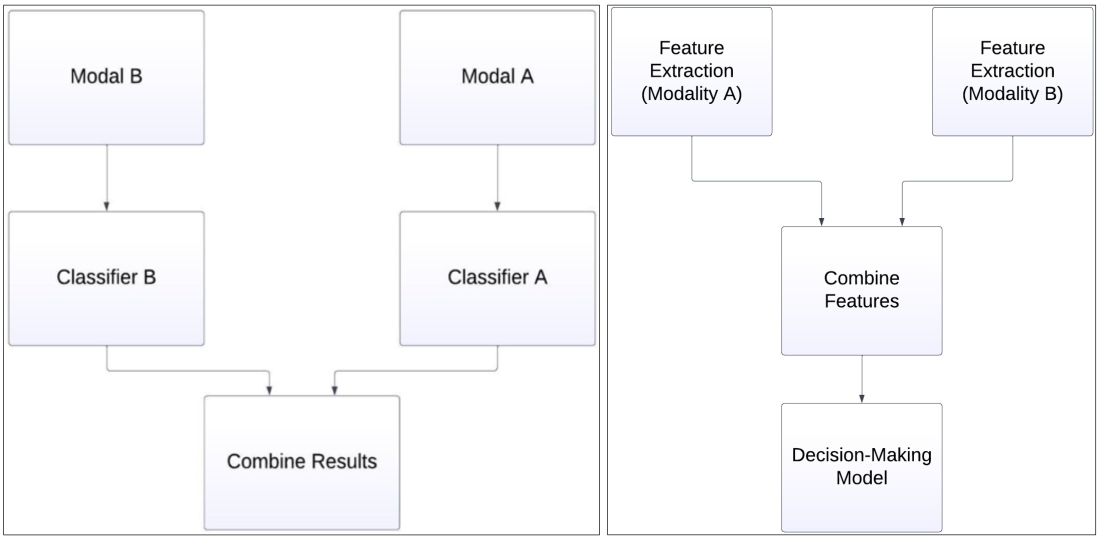


Figure 3. VGG16 architecture for feature extraction



(a) Late fusion

(b) Early fusion

Figure 4. Fusion architecture

This study leverages early fusion, a technique that integrates features from different modalities at the early stages of the processing workflow. In our case, this involves combining sensor data features (e.g., gas concentration) with image features extracted from thermal camera readings (e.g., thermal signature). This approach offers several advantages:

Potential for Improved Accuracy: By learning relationships between sensor data and thermal signatures, the model can potentially achieve more accurate gas location identification compared to separate processing of each modality [17].

Early fusion allows the model to compress a large set of variables from both sensors and thermal imaging into a smaller, more manageable feature vector. This simplifies the training process and can improve model performance, especially when dealing with limited training data [18].

The combined feature vector facilitates the model's ability to learn latent relationships between sensor data and thermal signatures that might not be evident when processed independently. This can lead to a more robust and generalizable model for gas classification tasks.

The combined features provide a richer representation of the data, potentially leading to better classification or recognition performance.

Early fusion requires training only one model on the combined feature vector. This approach simplifies the training process compared to late fusion, which necessitates training separate models for each modality and then combining their decisions.

t-Distributed Stochastic Neighbor Embedding (t-SNE) and Principal Component Analysis (PCA) visualizations were utilized to explore the distribution of features in a lower-dimensional space. These visualizations provided insights into the separability of different classes in the feature space. t-SNE helped in understanding the complex relationships between features and revealed clusters corresponding to different gas classes. Figures 5(a) and 5(b) give the PCA and t-SNE visualizations.

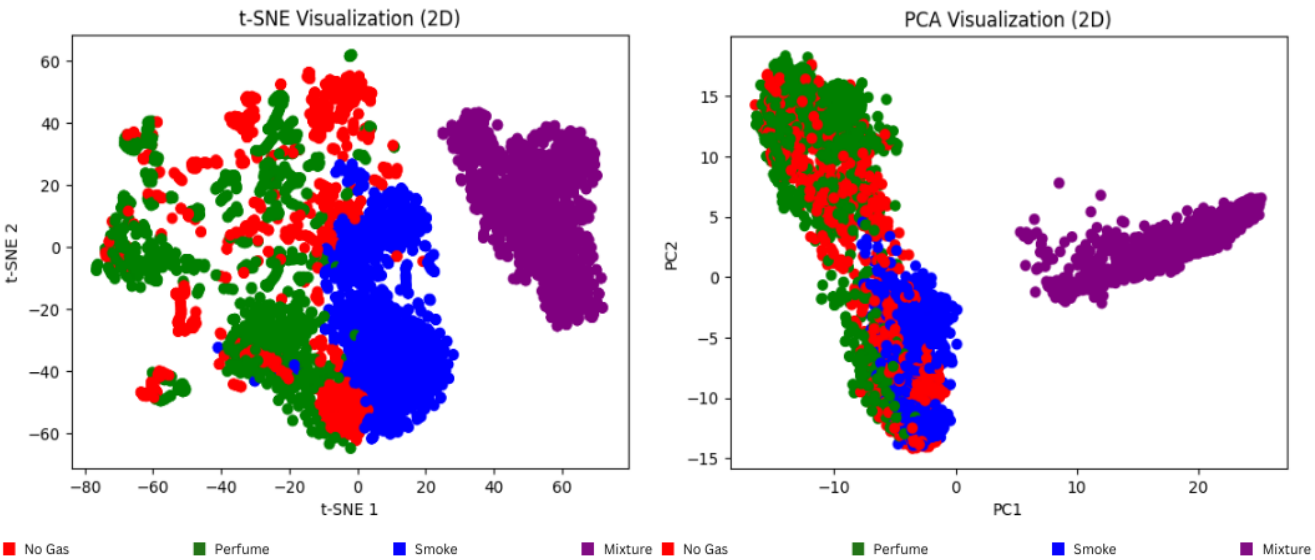
4.4 Models

The study involves comparison of three models, SVM (Support Vector Machine), RFC (Random Forest Classifier), and KNN (K Nearest Neighbor). This section elaborates the use of these models in multimodal classification after early fusion. The generic model architecture can be referenced from Figure 6.

4.4.1 Support Vector Machine (SVM)

SVMs are supervised learning models, mostly used for classification of data. They work by finding a hyperplane in high-dimensional space that most accurately separates data points belonging to different classes. This hyperplane maximizes the distance between the classes, essentially creating a clear decision boundary [10].

SVM is a highly trained model that effectively handles the feature vectors that are combined from various modalities like text and images, as used in this study. It also allows analysis of high-dimension feature space, which allows better identification of subtle relationship across modalities [19]. SVMs focus on the support vectors, leading to a sparse model that is less prone to overfitting, a common challenge in high-dimensional settings [20]. They also utilize kernel functions to project the data onto a high-dimensional space where classes become more separable, even when they are not linearly separable in the original feature space [20]. This is particularly beneficial for multimodal data that may exhibit complex relationships between modalities.



(a) t-SNE visualization (b) PCA visualization

Figure 5. Visualization post early fusion

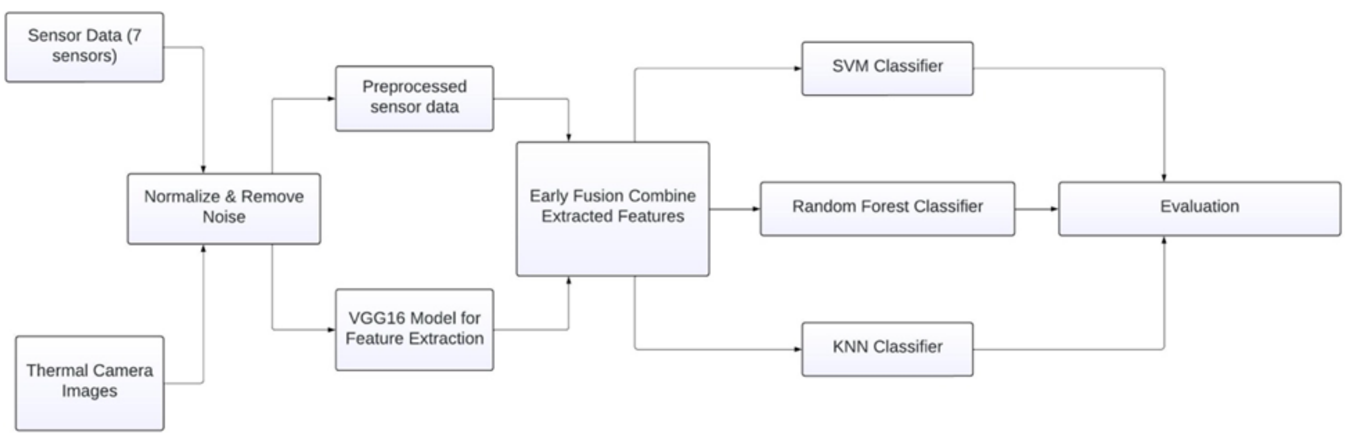


Figure 6. Generic model architecture

4.4.2 Random Forest Classifier

Random Forest Classifier (RFC) is a machine learning model based on supervised ensemble learning. It consists of multiple decision trees. Each tree is trained on a random subset of features and training data. This injects randomness into the learning process, helping to reduce overfitting. The final classification is made by combining the predictions from all individual trees [21].

RFC is particularly well-suited for this scenario due to its ability to handle effectively data that is high-dimensional [22]. The random subspace selection process during tree construction reduces the impact of irrelevant features from any modality, leading to robust predictions [21]. Additionally, RFC's feature importance scores can be used to identify the most influential features from each modality, aiding in data interpretation and model explainability [23].

4.4.3 K Nearest Neighbor

K-Nearest Neighbors (KNN) is a machine learning model used for classification and regression. This study utilized KNN for classification purposes, where the data points are classified using the labels of K Nearest Neighbors available. The data points are represented by feature vectors, and the distance metric (e.g., Euclidean distance). The distance shows how close the neighbors are. However, the performance of KNN heavily relies on choosing the optimal value of k [24].

The reason for choosing KNN in this study is because early fusion combines data from different modalities into one feature vector, as discussed before. KNN also allows understanding which neighbors contribute most to the classification, providing insights into the decision process, especially valuable in multimodal scenarios where reasoning across modalities is crucial [25]. Studies have shown that KNN can achieve competitive performance on multimodal classification tasks, particularly when the modalities share underlying structures or complementary information [26].

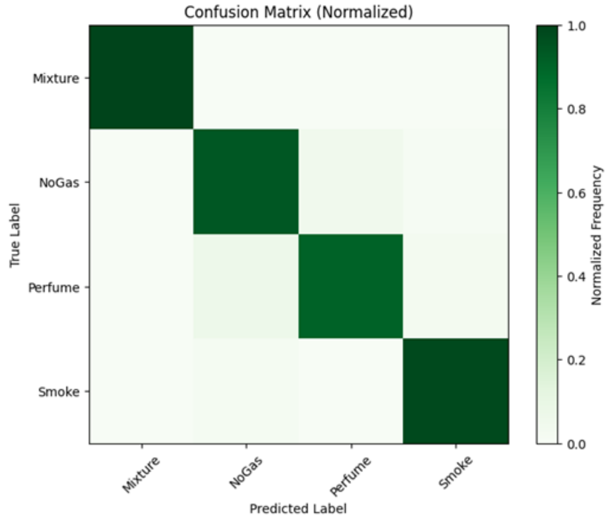
5. RESULTS

This section analyses the key findings obtained after the experimentation and compares the performance metrics of the employed machine learning models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest.

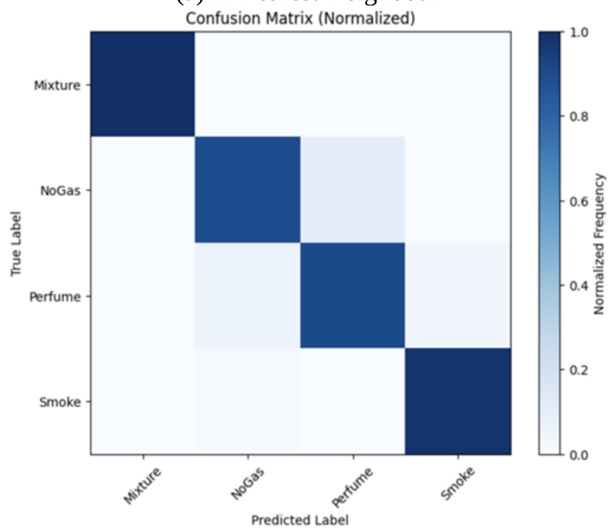
Our evaluation based on validation accuracy revealed a strong performance across all three machine learning models. Validation accuracy was selected as the primary metric for model evaluation. This metric directly reflects the model's ability to correctly classify unseen gas samples, which is crucial for real-world applications where accurate gas detection is essential. While other metrics like precision, recall, and F1-score provide valuable insights, validation accuracy offers a clear and concise measure of overall classification performance. Additionally, the confusion matrix presented in the results section provides a more detailed breakdown of the model's performance for each gas class.

The Random Forest model achieved superior performance, correctly classifying gas samples at a rate of 96%. Closely following were KNN and SVM models, both achieving an impressive accuracy of 94%.

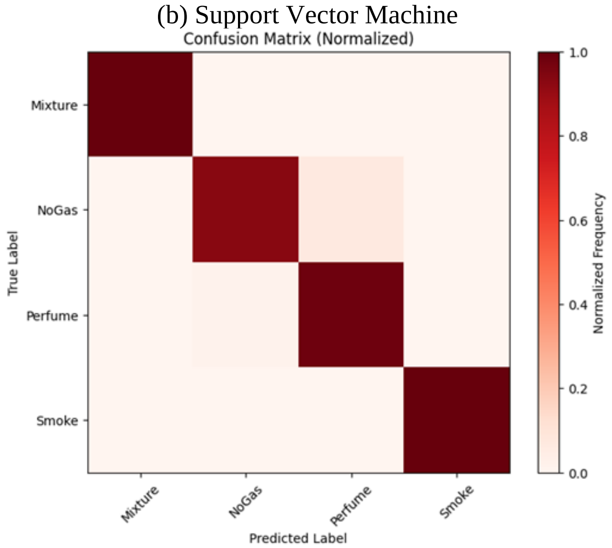
For achieving a more detailed view of the classification performance, Figure 7 presents the confusion matrices for all three machine learning models (KNN in Figure 7(a), SVM in Figure 7(b) and Random Forest in Figure 7(c)). Confusion matrices offer valuable insights into the true positive, false positive, true negative, and false negative rates for each gas class ("No Gas," "Smoke," "Perfume," and "Mixture"). By analyzing these matrices, we can gain a deeper understanding of any classification errors made by the models and identify potential areas for improvement. Table 1, gives the classification reports of each model, containing information about support, f1-score as well as recall and precision.



(a) K Nearest Neighbour



(b) Support Vector Machine



(c) Random Forest algorithms

Figure 7. Confusion matrices

Table 1. Performance metrics (accuracy, precision, recall, F1-score) of K-Nearest Neighbour, Support Vector Machine, Random Forest

Model	Accuracy	Precision	Recall	F1
KNN	0.9453	0.9452	0.9453	0.9452
SVM	0.9422	0.9449	0.9422	0.9423
RFC	0.9641	0.9676	0.9641	0.9642

The VGG16 architecture employed for feature extraction demonstrated strong performance in capturing relevant features from both the gas sensor and thermal image datasets. Its deep architecture allows for hierarchical feature extraction, enabling the model to learn complex patterns.

Early fusion, as demonstrated in this study, stands out as a robust fusion technique due to its ability to integrate information from multiple modalities at an early stage of the model architecture. By combining feature representations from both gas sensor readings and thermal images before feeding them into the machine learning models, early fusion enables the model to learn more comprehensive representations of the data. Moreover, early fusion typically reduces computational overhead compared to late fusion techniques, as it avoids the need for separate processing streams for each modality.

This study demonstrates the effectiveness of early fusion for gas classification. However, limitations exist. Current research utilizes a single feature extraction method (VGG16) for both sensor data and thermal images. Exploring domain-specific feature extraction techniques tailored to sensor data might further improve performance. Additionally, investigating alternative early fusion techniques and hyperparameter optimization could lead to further accuracy gains. Future work could also explore combining early and late fusion approaches for potentially enhanced performance.

Random Forest is particularly effective in multi-modal gas detection. Several theoretical underpinnings can support this claim. One is their ensemble learning nature. Random Forests do not rely on a single decision tree, but rather build a "forest" of them. A random subset of features is used by each tree at each split point, reducing the chances of over-fitting.

Random Forests also uses out-of-bag data for error estimation, which strengthens its performance. The model trains on a subset of the data and uses the remaining portion (out-of-bag data) to make performance assessments. This continuous evaluation helps the forest learn robust decision boundaries in the combined feature space.

While the overall performance of the Random Forest model aligns with expectations, a closer examination of the confusion matrix (Figure 7) reveals a minor anomaly. Specifically, the model exhibits a higher false positive rate for "Perfume" compared to "Smoke" (0.4 vs. 0.2). This is counterintuitive, as thermal signatures of smoke are generally expected to be more distinct from the background compared to perfume.

There are a few potential explanations for this observation. One possibility is that certain types of smoke, perhaps those with lower temperatures or similar burning materials used during data collection, might have led to thermal image features that partially overlap with those of perfume. Additionally, there might have been borderline cases where smoke traces lingered in the air, influencing the thermal signature and causing some confusion for the model during data collection. By addressing these potential causes of the minor deviation observed in the confusion matrix, we can refine the multi-sensor data fusion approach and further enhance the accuracy of gas classification.

6. CONCLUSION

In conclusion, this research explored early fusion for gas classification using a multimodal dataset with sensor data and thermal images. All three machine learning models achieved high validation accuracy (Random Forest: 96%, K-Nearest Neighbors and Support Vector Machine: 94%), demonstrating the potential of this approach for real-world applications. Although, the study can delve into optimizing fusion techniques, feature extraction, and potentially explore combining early and late fusion strategies for further performance gains. This research paves the way for the development of improved gas detection systems utilizing the rich information offered by multimodal data.

This research contributes to the growing body of knowledge on sensor data fusion for gas classification. While existing theoretical frameworks have laid the foundation for this field, limitations remain, such as the dependence on specific feature engineering techniques or restricted sensor modalities. Our work with early fusion and the combination of MOS sensors with thermal image features demonstrates the potential for improved gas classification accuracy by leveraging the complementary strengths of each modality. These findings not only validate the effectiveness of this approach but also suggest avenues for advancing existing theories. It paves the way for the development of more generalizable frameworks applicable to a broader range of gas detection tasks. Furthermore, our research opens doors for further theoretical exploration. Future work could investigate the integration of additional sensor modalities, such as gas chromatography or spectroscopic techniques, to potentially achieve even more comprehensive gas classification. Additionally, exploring deep learning-based fusion techniques could be a promising avenue for further research.

7. FUTURE SCOPE

This study demonstrates the effectiveness of multi-sensor data fusion for enhanced gas classification performance. However, there are several areas for further exploration:

- Expanding the Sensor Suite: The current study utilizes data from seven sensors. Investigating the impact of including additional sensors with different functionalities (e.g., catalytic bead, photoionization detector) could provide a richer feature space and potentially improve classification accuracy and generalizability.
- Exploration of Advanced Fusion Techniques: This study employs early fusion for combining sensor data and image features. Investigating the effectiveness of other fusion techniques, such as mid-level or late fusion, could be explored to determine if these approaches offer further

performance gains.

- Real-Time Implementation: While this study focused on offline classification, developing a real-time gas detection system using the proposed approach would be a valuable next step. This would involve addressing computational efficiency and exploring online learning algorithms for adapting to dynamic environments.
- Gas Concentration Estimation: The current model focuses on gas classification. Extending the model to estimate gas concentration levels would provide valuable information for risk assessment and response protocols.
- Testing with Diverse Gas Mixtures: The current dataset includes four classes. Testing the model's performance with a broader range of gas mixtures, especially those encountered in real-world applications, would enhance the generalizability and robustness of the approach.
- Computational Efficiency: While the current study achieved high accuracy, the computational cost of processing data from seven sensors and thermal images might limit real-time implementation. Future work could explore techniques for feature selection and dimensionality reduction to optimize the model for real-time applications without compromising accuracy.

ACKNOWLEDGMENT

This work is funded by Universidad Nacional Mayor de San Marcos (UNMSM) according RR N° 013865-2021-R/UNMSM. Also, we acknowledge the Advanced and Innovative Research Laboratory (AAIR Labs: www.aairlab.com) India for technical support.

REFERENCES

[1] Hasan, F. (2024). Alarming rise in fire incidents in India indicates authorities' negligence, safety lapses. EducationPost. https://educationpost.in/news/national/alarming-rise-fire-incidents-india-indicates-authorities-negligence-safety-lapses.

[2] Pardo, M., Sberveglieri, G. (2005). Classification of electronic nose data with support vector machines. Sensors and Actuators B: Chemical, 107(2): 730-737. https://doi.org/10.1016/j.snb.2004.12.005

[3] Pardo, M., Sberveglieri, G. (2005). Classification of electronic nose data with support vector machines. Sensors and Actuators B: Chemical, 107(2): 730-737. https://doi.org/10.1016/j.snb.2004.12.005

[4] Peng, P., Zhao, X.J., Pan, X.F., Ye, W.B. (2018). Gas classification using deep convolutional neural networks. Sensors, 18(1): 157. https://doi.org/10.3390/s18010157

[5] Oh, J., Hwang, H., Nam, Y., Lee, M.I., Lee, M.J., Ku, W., Song, H.W., Pouri, S.S., Lee, J.O., An, K.S., Yoon, Y., Lim, J., Hwang, J.H. (2022). Machine learning-assisted gas-specific fingerprint detection/classification strategy based on mutually interactive features of semiconductor gas sensor arrays. Electronics, 11(23): 3884. https://doi.org/10.3390/electronics11233884

[6] Adefila, K., Yan, Y., Wang, T. (2015). Leakage detection of gaseous CO_2 through thermal imaging. In 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, Pisa, Italy, pp. 261-265. https://doi.org/10.1109/I2MTC.2015.7151276

[7] Xiong, K., Jiang, J., Pan, Y., Yang, Y., Chen, X., Yu, Z. (2022). Deep learning approach for detection of underground natural gas micro-leakage using infrared thermal images. Sensors, 22(14): 5322. https://doi.org/10.3390/s22145322

[8] Chen, M., Xia, X.W., Kang, Z., Li, Z.N., Dai, J.M., Wu, J.Y., Chen, C., Qiu, Y., Liu, T., Liu, Y.X., Zhang, Z.Y., Shen, Q.N., Tao, S.C., Deng, Z.X., Lin, Y., Wei, Q.L. (2024). Distinguishing schizophrenia and bipolar disorder through a multiclass classification model based on multimodal neuroimaging data. Journal of Psychiatric Research, 172: 119-128. https://doi.org/10.1016/j.jpsychires.2024.02.024

[9] Dyrba, M., Grothe, M., Kirste, T., Teipel, S.J. (2015). Multimodal analysis of functional and structural disconnection in a lzheimer's disease using multiple kernel SVM. Human Brain Mapping, 36(6): 2118-2131. https://doi.org/10.1002/hbm.22759

[10] Akula, A., Sardana, H.K. (2019). Deep CNN-based feature extractor for target recognition in thermal images. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON), Kochi, India, pp. 2370-2375. https://doi.org/10.1109/TENCON.2019.8929697

[11] Rahate, A., Mandaokar, S., Chandel, P., Walambe, R., Ramanna, S., Kotecha, K. (2023). Employing multimodal co-learning to evaluate the robustness of sensor fusion for industry 5.0 tasks. Soft Computing, 27(7): 4139-4155. https://doi.org/10.1007/s00500-022-06802-9

[12] Narkhede, P., Walambe, R., Mandaokar, S., Chandel, P., Kotecha, K., Ghinea, G. (2021). Gas detection and identification using multimodal artificial intelligence based sensor fusion. Applied System Innovation, 4(1): 3. https://doi.org/10.3390/asi4010003

[13] Attallah, O. (2023). Multitask deep learning-based pipeline for gas leakage detection via E-nose and thermal imaging multimodal fusion. Chemosensors, 11(7): 364. https://doi.org/10.3390/chemosensors11070364

[14] Dwivedi, S., Goel, T., Tanveer, M., Murugan, R., Sharma, R. (2022). Multimodal fusion-based deep learning network for effective diagnosis of Alzheimer's disease. IEEE MultiMedia, 29(2): 45-55. https://doi.org/10.1109/MMUL.2022.3156471

[15] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.

[16] Asskali, S. (2017). Polyp detection: Effect of early and late feature fusion. Doctoral Dissertation, University of Oslo.

[17] Huhtanen, J.T., Nyman, M., Doncenco, D., Hamedian, M., Kawalya, D., Salminen, L., Sequeiros, R.B., Koskinen, S.K., Pudas, T.K., Kajander, S., Niemi, P., Hirvonen, J., Aronen, H.J., Jafaritadi, M. (2022). Deep learning accurately classifies elbow joint effusion in adult and pediatric radiographs. Scientific Reports, 12(1): 11803. https://doi.org/10.1038/s41598-022-16154-x

[18] Gunes, H., Piccardi, M. (2005). Affect recognition from face and body: early fusion vs. late fusion. In 2005 IEEE international conference on systems, man and cybernetics, Waikoloa, HI, USA, Vol. 4, pp. 3437-3443. https://doi.org/10.1109/ICSMC.2005.1571679

[19] El-Sappagh, S., Saleh, H., Sahal, R., Abuhmed, T., Islam, S.R., Ali, F., Amer, E. (2021). Alzheimer's disease

progression detection model based on an early fusion of cost-effective multimodal data. Future Generation Computer Systems, 115: 680-699. https://doi.org/10.1016/j.future.2020.10.005

[20] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning, 20: 273-297. https://doi.org/10.1007/BF00994018

[21] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning. New York: Springer. https://doi.org/10.1007/978-1-0716-1418-1

[22] Paris, R.B. (2020). Uniform asymptotics of a Gauss hypergeometric function with two large parameters, V. arXiv Preprint, arXiv:2004.01945. https://doi.org/10.48550/arXiv.2004.01945

[23] Belgiu, M., Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. Journal of Photogrammetry and Remote Sensing, 114: 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[24] Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN model-based approach in classification. In on the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE (OTM 2003), Catania, Sicily, Italy, pp. 986-996. https://doi.org/10.1007/978-3-540-39964-3_62

[25] Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., de Albuquerque, V.H.C. (2018). Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cognitive Systems Research, 52: 36-48. https://doi.org/10.1016/j.cogsys.2018.06.006

[26] Lamba, R., Gulati, T., Alharbi, H.F., Jain, A. (2022). A hybrid system for Parkinson's disease diagnosis using machine learning techniques. International Journal of Speech Technology, 25: 583-593 https://doi.org/10.1007/s10772-021-09837-9