Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 3 (2025)

Enhanced Machine Learning Framework for Chronic Gastritis Prediction from Electronic Nose Breath Sensor Data

Poornima Eswaran[1*], Chandra Eswaran[2]
[1,2]Department of computer Science, Bharathiar University, Tamil Nadu, India
Corresponding Author: Poornima Eswaran

Abstract: - Chronic gastritis is a prevalent inflammatory condition of the gastric mucosa that, if not properly diagnosed, it can progress to ulcers or gastric cancer. Breath analysis using non-invasive electronic nose (e-nose) systems has developed as a promising diagnostic approach, detecting volatile organic compounds associated with the disease. This paper presents a hybrid Fuzzy Deep Convolutional Neural Network (F-DCNN) framework for gastritis prediction from e-nose sensor data. The proposed method integrates pre-processing, feature selection, deep convolutional feature extraction, and fuzzy inference to identify individual patterns. Experiments comparing SVM, k-NN, ANN, conventional DCNN, and F-DCNN show that F-DCNN achieves superior performance, with an accuracy of 96.23% and the highest AUC among all tested classifiers. The proposed F-DCNN algorithm achieves efficient processing, ensuring fast execution. These results highlight the potential of hybrid deep-fuzzy models in medical diagnostics, offering both high accuracy and interpretability.

Keywords: *Non-invasive, Breath data, Chronic Gastritis, Feature Selection, Machine Learning*

1. Introduction

Chronic Gastritis, an inflammation of the gastric mucosa, is a prevalent gastrointestinal disorder with global health associations. Its causes include *Helicobacter pylori* infection, chronic use of nonsteroidal anti-inflammatory drugs (NSAIDs), extreme alcohol consumption, and autoimmune processes [1]. Symptoms such as epigastric discomfort, nausea, and bloating are often non-specific, and in many cases, the condition is asymptomatic until complications such as peptic ulcer disease or, in chronic cases, gastric carcinoma arise. Early and accurate detection is therefore critical for effective intervention [2]. Upper gastrointestinal endoscopy, combined with histopathological examination, remains the traditional method for diagnosing gastritis, offering high diagnostic accuracy but it is invasive, and often uncomfortable for patients [3]. Other non-invasive techniques, like stool antigen assays, urea breath tests, and serological testing, are either indirect, exclusive to gastritis caused by H. pylori, or impacted by recent medication usage. These drawbacks have increased interest in innovative, quick, precise, and patient-friendly non-invasive diagnostic techniques [4].

Exhaled human breath contains a complex mixture of volatile organic compounds (VOCs) rising from metabolic processes and microbial activity. In gastric diseases, inflammation and infection can alter VOC composition manifesting as changes in compounds such as ammonia, ethanol, and various aldehydes. Electronic nose (e-nose) systems, which use arrays of cross-reactive gas sensors, have emerged as a promising technology for detecting these VOC patterns [5].

Advances in machine learning (ML) provide great approach for understanding high-dimensional sensor data. Through pattern recognition, ML algorithms can uncover indirect relationships between VOC profiles and disease states, enabling accurate classification even in the presence of biological variability and environmental noise. A critical step in this process is feature selection, which includes identifying the most relevant and discriminative sensor-derived variables from the often large and noisy dataset. Effective feature selection reduces dimensionality, overfitting, and improves model interpretability, thereby enhancing diagnostic performance.Once relevant

features are identified, and classification algorithms are applied to assign each sample to a diagnostic category such as gastritis or non-gastritis based on learned patterns in the data. Evaluating multiple classifiers allows for selecting the approach that offers the best balance of accuracy, sensitivity, and specificity for the task [6].

In this work, an existing publicly available e-nose breath dataset to examine the possibility of predicting gastritis in a non-invasive manner. In this approach involves three main stages: (1) data pre-processing to enhance signal quality and remove noise; (2) feature selection to identify the most relevant features; and (3) classification using machine learning algorithms to distinguish between gastritis and non-gastritis.

2. Related work

Breath analysis diagnostics have been explored widely for various medical conditions, with research across non-invasive sensing technologies, feature selection techniques, and machine learning classification models. Previous studies have proved that exhaled VOC patterns can serve as reliable biomarkers for diseases such as lung cancer, COPD, diabetes, and gastric cancer, using sensor arrays coupled with statistical or machine learning analysis. Feature selection methods [7] such as PCA, ReliefF, and LASSO have been shown to enhance model performance by reducing dimensionality and improving interpretability, while classifiers like SVM, Random Forest, and Gradient Boosting consistently achieve high diagnostic accuracy. Table 1 summarizes representative research across these areas, providing a comparative view of their objectives, methodologies, and key findings.

Table 1. List of Related work

Author(s)	Year	Focus	Results
Rangel et al. [8]	2025	H. pylori Breath Test	MIR spectroscopic breath test, high accuracy, non-invasive, point-of-care potential
Haick et al. [9]	2013	Cancer Detection via VOCs	Nanoparticle e-nose; high sensitivity/specificity for stomach cancer
Wang et al. [10]	2009	Laser Spectroscopy for Breath Analysis	Reviewed laser-based VOC detection; identified gastric biomarkers
Lawal et al. [11]	2017	VOC Analysis Methods	Surveyed GC-MS and sensor arrays; disease-specific VOC signatures
De Vries et al. [12]	2015	E-Nose + Spirometry	Multimodal integration improved respiratory diagnostics
Mule & Patil [13]	2021	Statistical Methods	Preprocessing, feature extraction, ML for breath data; improved detection accuracy
Meister et al. [14]	2021	Audio Feature Ranking	Feature selection improved COVID-19 detection by 17%
Xia et al. [15]	2022	Audio-Based Respiratory Screening	MFCC and spectral features for respiratory condition detection
Ezzat et al. [16]	2021	Diabetes via Breath VOCs	ReliefF feature selection; >90% classification accuracy
Yang et al. [17]	2022	VOC-Based Diabetes Detection	SVM/KNN classifiers; >90% sensitivity and specificity
Wei et al. [18]	2017	Lung Sound Classification	CNN and SVM; ~86% accuracy
Alshammari et al. [19]	2023	COVID-19 via E-Nose VOCs	Gradient boosting; ~96% accuracy
Zhang et al. [20]	2024	Lung Cancer Breath Test	SVM + SMOTE; highest accuracy on chemosensor datasets

Rahman et al. [21]	2020	CNN-RNN Respiratory Classification	Hybrid deep learning; ~71.8% accuracy
Li et al. [22]	2024	mmWave Radar Breath Rates	Quadratic SVM; 95% accuracy
Xu et al. [23]	2013	Gastric Disease VOC Differentiation	Nanomaterial sensors; 89% sensitivity, 90% specificity
Amal et al. [24]	2016	Nanoarray & GC-MS Gastric Cancer	Up to 98% specificity, 97% sensitivity
Schuermans et al. [25]	2020	E-Nose Gastric Cancer	MOS e-nose; 81% sensitivity, 71% specificity
Amal et al. [26]	2013	Early vs Advanced Gastric Cancer	VOC profiling; 89% sensitivity, 94% specificity

3. Methodology

3.1 Pre-processing

The dataset used in this study was obtained (https://github.com/Chenyif/enose) [27] which contains e-nose measurements from two groups: healthy individuals and patients diagnosed with chronic gastritis. Each record represented the time-series output of multiple metal-oxide semiconductor (MOS) sensors for a single subject.

Missing Values: The raw sensor data contained missing readings produced by sensor instability. Missing values within short gaps were imputed using linear interpolation between the nearest valid samples. If ta and tb represent the timestamps before and after a missing point tm, the imputed value was calculated as:

$$x(t_m) = x(t_a) + \frac{t_m - t_a}{t_b - t_a} . [\, x(t_b) - x(t_b)] \quad (1)$$

Baseline Correction: Electronic nose sensors often exhibit baseline drift due to environmental factors or sensor drift. To normalize the signal, the mean pre-exposure value bi of each sensor i was computed and subtracted from the corresponding time-series readings:

$$x_i'(t) = x_i(t) - b_i \quad (2)$$

This ensured that all responses were aligned to a common zero reference point, enabling more accurate feature comparison.

Noise Reduction: High-frequency noise rising from sensor electronics and ambient conflicts was minimized using a moving average filter. For each time point t, the smoothed value was obtained as:

$$\tilde{x}(t) = \frac{1}{w} \sum_{k=0}^{w-1} x(t - k) \quad (3)$$

Where w is the smoothing window size.

Label Encoding: The dataset consisted of two groups healthy subjects and patients with chronic gastritis. These categorical class labels were converted into numerical form for machine learning compatibility, with "0" representing healthy and "1" representing gastritis cases. This encoding allowed supervised learning algorithms to understand the target variable during training and evaluation.

3.2 Feature Selection

Feature selection is an essential step in machine learning approach, particularly for high-dimensional sensor data such as that produced by electronic noses. The goal is to retain only the most informative and non-redundant features, thereby improving classification performance, reducing overfitting, and lowering computational costs.

In this study, a hybrid approach was implemented, combining a filter method for statistical preselection with Principal Component Analysis (PCA) for variance-based dimensionality reduction. Filter methods, such as correlation analysis and mutual information ranking, for exploration because they quickly calculate the statistical connection between each feature and the target variable. However, these methods alone may retain redundant features or fail to capture latent patterns in the data. Instead, Principal Component Analysis (PCA) is a powerful exploitation technique that converts features and concentrating variance into a compact representation. When used in isolation, PCA may still embed irrelevant features, reducing interpretability and classifier efficiency. The proposed hybrid Filter–PCA approach leverages the strengths of both strategies: the filter stage first explores the entire feature set, removing irrelevant and highly correlated variables, while the PCA stage exploits the refined subset to produce an optimized, low-dimensional representation. This combination shows that the final feature set is both statistically relevant and information-dense, making it particularly effective for non-invasive gastritis prediction from breath sensor data.

3.2.1 Method 1: Filter-Based Feature Selection

Filter methods rank features according to statistical relevance to the target variable without relying on a specific classifier. In this work, two filtering strategies were applied sequentially:

Correlation Filtering

In the first stage of the hybrid feature selection process, correlation filtering was applied to detect and remove redundant features. The Pearson correlation coefficient was computed between every possible pair of features to measure the strength and direction of their linear relationship. If two features were found to be highly correlated ($|r|>0.9$), one of them was removed to reduce redundancy in the subsequent modeling stages. This filtering step confirmed that only features carrying unique and non-overlapping information were retained for further analysis. The Pearson correlation coefficient was calculated using Eq. (4).

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (4)$$

Mutual Information Ranking

After removing redundant features through correlation filtering, the remaining features were ranked based on their statistical dependency with the target variable using Mutual Information (MI). MI quantifies the amount of information one variable contains about another, making it suitable for identifying features that are most informative for classification. This step prioritizes variables that contribute the most to predicting the class labels. The top k features with the highest MI scores were retained for the next stage of dimensionality reduction. Mutual Information was calculated using Eq. (5).

$$MI(X, Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (5)$$

3.2.2 Method 2: Principal Component Analysis (PCA)

Following the filter-based pre-selection of features, PCA algorithm was applied to further reduce dimensionality while stabilising the maximum possible change in the data. PCA changes the selected features into a new set of principal components, which are ordered allowing to the amount of variance they explain [28]. The covariance matrix of the filtered features was first computed, and then its eigenvalues and eigenvectors were obtained. The principal components corresponding to the largest eigenvalues were selected such that the cumulative variance explained met or exceeded the predefined threshold (τ=95%). The covariance matrix was calculated as shown in Eq. (6).

$$C = \frac{1}{n-1} X^T X \quad (6)$$

The data were then projected onto the selected principal components using Eq. (7).

$$Z = X.W \quad (7)$$

Where X is the mean-centered feature matrix, W is the matrix of selected eigenvectors, and Z is the transformed dataset in the reduced feature space. The pseudo code of hybrid feature selection algorithm represent in Algorithm 1.

Algorithm 1: Hybrid Filter–PCA Feature Selection

Input:

X: Feature matrix (n samples × p features)

Y: Class labels

r_thresh: Correlation threshold

k: Number of top features after MI ranking

τ: PCA variance retention threshold

Output:

X_PCA: Reduced feature matrix

Steps:

1. Stage 1 – Exploration (Filter Selection)
2. For each feature pair (f_i, f_j) in X:
3. Compute Pearson correlation r_ij
4. If |r_ij| > r_thresh: Remove one of (f_i, f_j)
5. For each remaining feature f_i:
6. Compute Mutual Information MI(f_i, Y)
7. Rank features by MI in descending order
8. Select top k features → X_filtered
9. Stage 2 – Exploitation (PCA Transformation)
10. Center X_filtered (zero mean per feature)
11. Compute covariance matrix C of X_filtered
12. Compute eigenvalues λ_j and eigenvectors v_j of C
13. Sort eigenvectors in descending λ_j order
14. Select m components s.t. $(\Sigma_{j=1}^{m} \lambda_j)/(\Sigma_{j=1}^{p} \lambda_j) \geq \tau$
15. Transform: X_PCA = X_filtered × [v_1, v_2, ..., v_m]
16. Return X_PCA

3.3 Classification

The selected features obtained from pre-processing and feature selection are transform into a Deep Convolutional Neural Network (DCNN) designed to extract high-level, discriminative representations from the sensor data. The DCNN comprises multiple convolutional layers, each followed by nonlinear activation functions and pooling layers, which together capture complex spatial-temporal patterns inherent in the e-nose measurements. The output of the DCNN is a set of deep feature activations that summarize the raw input into a more informative feature space. To handle the uncertainty and variability present in sensor responses, these deep features are then transformed into fuzzy membership degrees across predefined linguistic categories such as Low, Medium, and High, using Gaussian membership functions. This fuzzification step allows the system to express the degree to which each feature belongs to these fuzzy sets, providing a soft and flexible representation of the data. Subsequently, a fuzzy inference system applies a rule base composed of if-then fuzzy rules that relate fuzzified features to class labels, combining antecedents with fuzzy logical operators such as minimum or product to calculate each rule's firing strength. The fuzzy outputs from all relevant rules are aggregated for each class and then defuzzified using the centroid method to produce crisp membership scores. The final predicted class is selected based on the highest defuzzified score, distinguishing between healthy and gastritis cases [29]. The DCNN is trained end-to-end via backpropagation to optimize convolutional filters and fully connected weights for accurate feature extraction. Concurrently or sequentially, the fuzzy membership functions and inference rules are tuned through optimization techniques to maximize classification performance. To enhance model efficiency and reduce overfitting, minimal weight quantification methods are employed on the DCNN parameters, which involve pruning or discretizing weights to lower precision representations while preserving predictive accuracy. This hybrid approach synergizes the powerful feature learning ability of DCNNs with the robust uncertainty modeling of fuzzy logic, resulting in an effective classifier for non-invasive gastritis prediction using e-nose data. Figure 2 shows the methodology of proposed Fuzzy DCNN classifier. The pseudo code of proposed classifier represent in Algorithm 2.

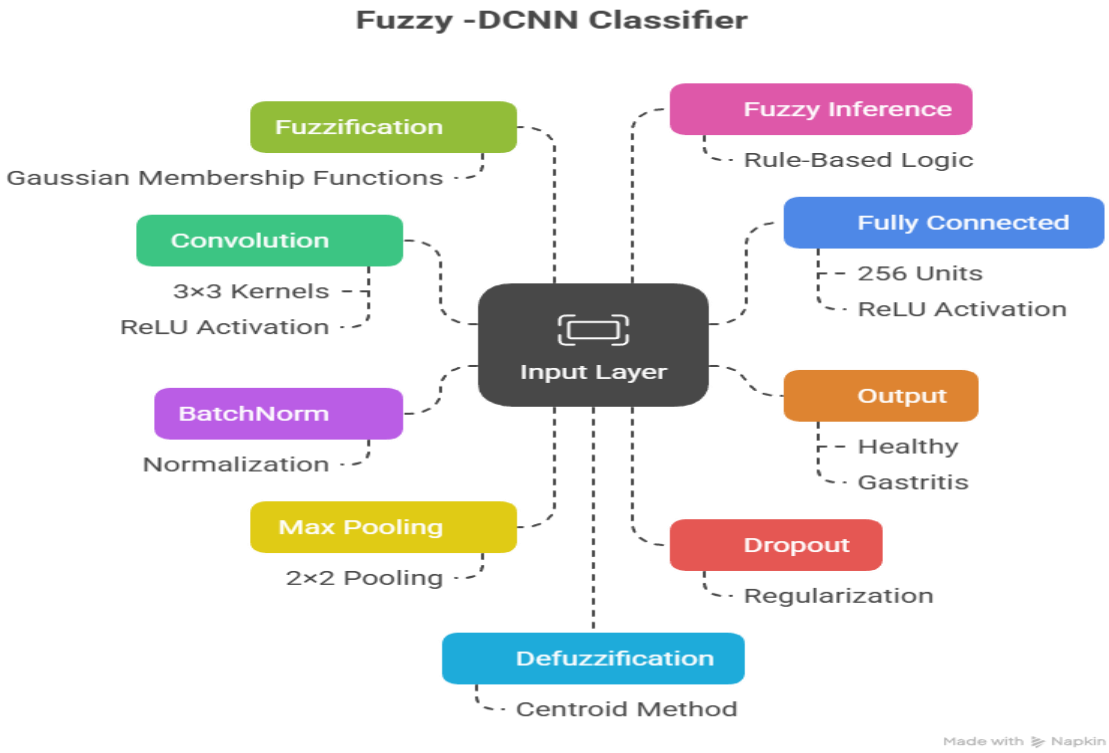

Fig. 1: Proposed F-DCNN Classifier Methodology

Algorithm 2: Proposed F-DCNN Classifier

Input:

```
    X_train, Y_train  # Training features and labels (post feature selection)
    X_test           # Test features
    fuzzy_params      # Initial fuzzy membership functions and rules
Output:
    Y_pred           # Predicted labels for X_test
```

Algorithm:

Step 1: Train DCNN

```
DCNN_model = initialize_dcnn()
DCNN_model.train(X_train, Y_train)
```

Step 2: Extract features from trained DCNN

```
def extract_features(model, X):
    features = model.get_activations(X, layer='final_layer')
    return features
train_features = extract_features(DCNN_model, X_train)
test_features = extract_features(DCNN_model, X_test)
```

Step 3: Fuzzy features

```
def fuzzify(features, fuzzy_params):
    fuzzified = []
    for feature_vector in features:
        fuzzified_vector = []
        for i, val in enumerate(feature_vector):
            memberships = compute_membership(val, fuzzy_params[i])
            fuzzified_vector.append(memberships)
        fuzzified.append(fuzzified_vector)
    return fuzzified
train_fuzzy = fuzzify(train_features, fuzzy_params)
test_fuzzy = fuzzify(test_features, fuzzy_params)
```

Step 4: Fuzzy inference and defuzzification

```
def fuzzy_inference(fuzzified_vector, fuzzy_rules):
    rule_outputs = []
    for rule in fuzzy_rules:
        antecedent_degree = evaluate_rule_antecedent(fuzzified_vector, rule)
        consequent_degree = rule.consequent * antecedent_degree
        rule_outputs.append(consequent_degree)
    aggregated = aggregate_rules(rule_outputs)
```

```
    crisp_score = defuzzify(aggregated)
    return crisp_score
Step 5: Predict labels for test data
Y_pred = []
for fuzz_vec in test_fuzzy:
    scores = fuzzy_inference(fuzz_vec, fuzzy_rules)
    predicted_label = argmax(scores)
    Y_pred.append(predicted_label)
return Y_pred
```

Table 2 presents the detailed structure of the proposed fuzzy-DCNN architecture designed for non-invasive gastritis prediction. The network begins with an input layer that takes the selected features from preprocessing. It then passes through a series of convolutional layers with small kernel sizes (3×3), each followed by batch normalization and ReLU activation to enhance feature extraction and improve training stability. Max pooling layers reduce spatial dimensions, helping to capture hierarchical patterns efficiently. The convolutional blocks are followed by fully connected layers that condense the learned features into a compact representation. A dropout layer is included to prevent overfitting. The final fully connected layer outputs features that are then fuzzified using Gaussian membership functions to model uncertainty. The fuzzified features are processed through a fuzzy inference system implementing rule-based logic, and finally, the fuzzy outputs are defuzzified to produce crisp class predictions distinguishing healthy individuals from gastritis patients. This layered architecture synergizes deep learning's feature extraction capabilities with fuzzy logic's handling of uncertainty to improve classification accuracy.

Table 2: Description of DCNN Layers

Layer No.	Layer Type	Output Shape	Kernel Size / Units	Activation	Remarks
1	Input Layer	(Input feature size)	–	–	Input from selected features
2	Convolutional Layer	(…, 64)	3×3	ReLU	Feature extraction
3	Batch Normalization	(…, 64)	–	–	Stabilizes training
4	Max Pooling Layer	(…, 32)	2×2	–	Downsampling
5	Convolutional Layer	(…, 128)	3×3	ReLU	Deeper feature extraction
6	Batch Normalization	(…, 128)	–	–	–
7	Max Pooling Layer	(…, 64)	2×2	–	–
8	Fully Connected Layer	256	–	ReLU	High-level feature representation
9	Dropout	256	–	–	Prevents overfitting
10	Fully Connected Layer	Number of fuzzy sets (e.g., 3)	–	Linear	Features for fuzzification

11	Fuzzification Module	(Same as above)	–	–	Membership functions applied
12	Fuzzy Inference Layer	Number of classes (2)	–	–	Applies fuzzy rules and aggregates
13	Defuzzification Layer	Number of classes (2)	–	–	Converts fuzzy outputs to crisp prediction

4. Experimental Results

The pre-processed and selected features e-nose dataset was fed into multiple machine learning classifiers, including k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Artificial Neural Network (ANN), standard Deep Convolutional Neural Network (DCNN), and the proposed Fuzzy DCNN approach. Performance evaluation was carried out using five key metrics as Accuracy, Precision, Recall, F1-score, and AUC under a 10-fold cross-validation scheme to ensure robustness and reduce bias in the results.

The confusion matrix for each classifier was computed to assess their capability for accurate classification of healthy and gastritis cases. The comparative diagnostic performance of these classifiers is summarized in Table 3.

Table 3: Comparison of Performance Measures

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC
Support Vector Machine (SVM)	89.15	87.40	88.23	87.25	89.50	90.25
k-Nearest Neighbors (kNN)	90.23	90.58	91.75	89.89	91.68	90.36
Artificial Neural Network (ANN)	91.89	91.36	92.78	91.54	92.36	90.89
DCNN	93.25	92.36	94.21	93.69	94.56	93.98
Proposed F-DCNN	96.23	95.89	96.47	94.92	96.32	95.98

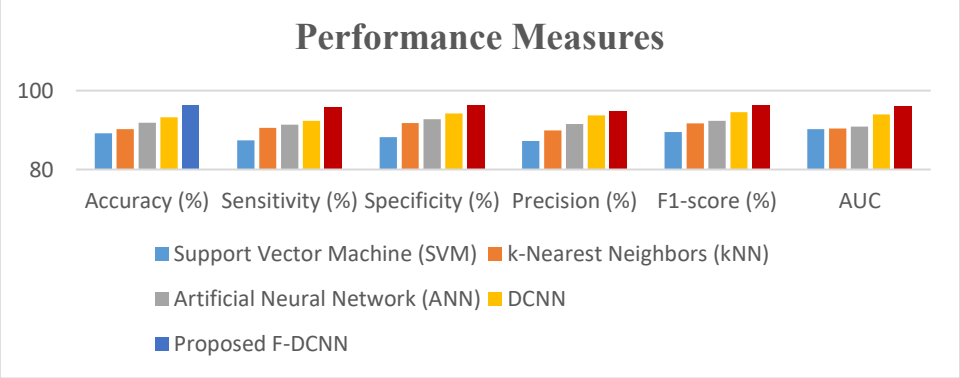

Fig. 2: Performance comparison of the proposed F-DCNN with existing classifiers.

The proposed F-DCNN achieves the highest accuracy, sensitivity, specificity, precision, F1-score, and AUC, demonstrating superior robustness and diagnostic capability for non-invasive gastritis prediction.

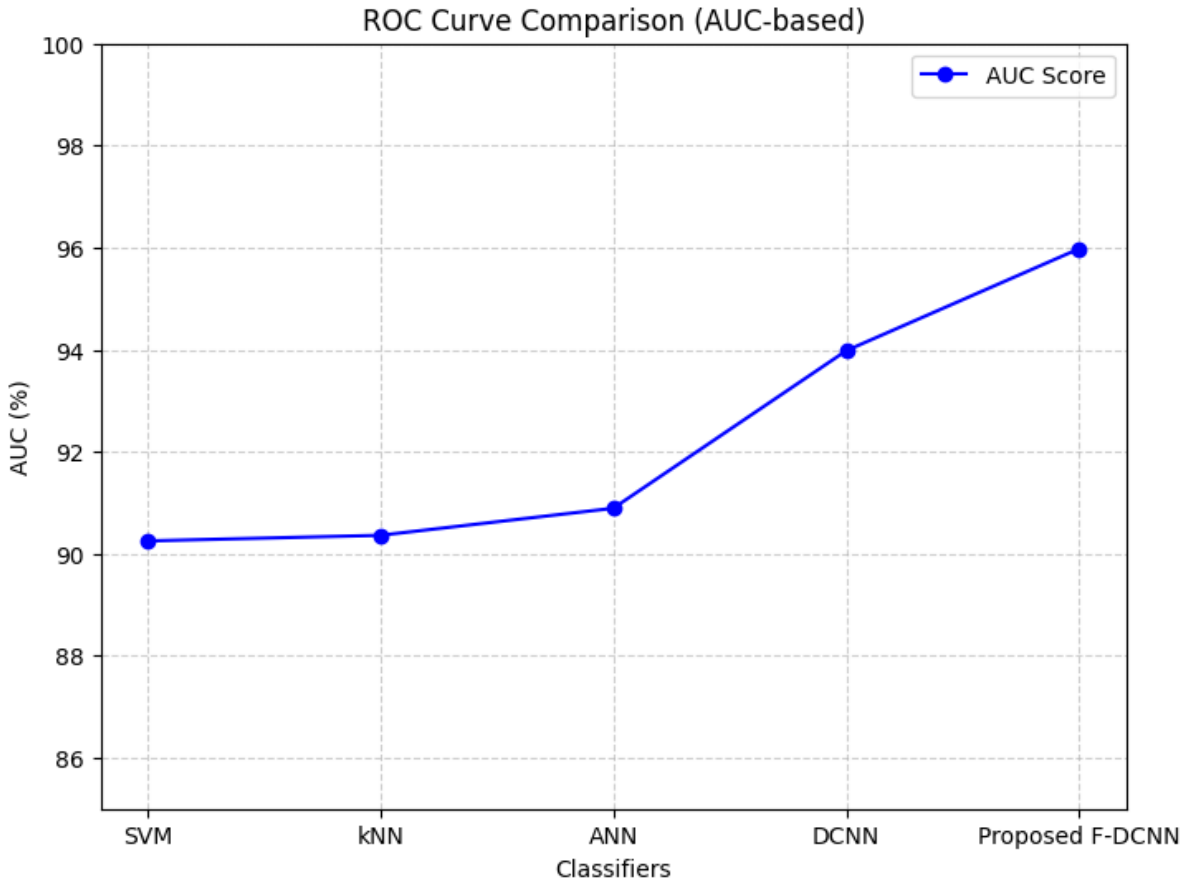

Fig. 3: Receiver Operating Characteristic (ROC) curve of the proposed F-DCNN model.

The curve illustrates the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across various classification thresholds. The high Area Under the Curve (AUC = 95.98%) indicates the model's strong discriminative capability in distinguishing between healthy subjects and gastritis cases.

Table 4: Comparative accuracy performance of classifiers before and after feature selection

Classifier	Accuracy (%) Before feature Selection	Accuracy (%) After feature Selection
Support Vector Machine (SVM)	85.05	89.15
k-Nearest Neighbors (kNN)	86.23	90.23
Artificial Neural Network (ANN)	87.36	91.89
DCNN	88.56	93.25
Proposed F-DCNN	91.08	96.23

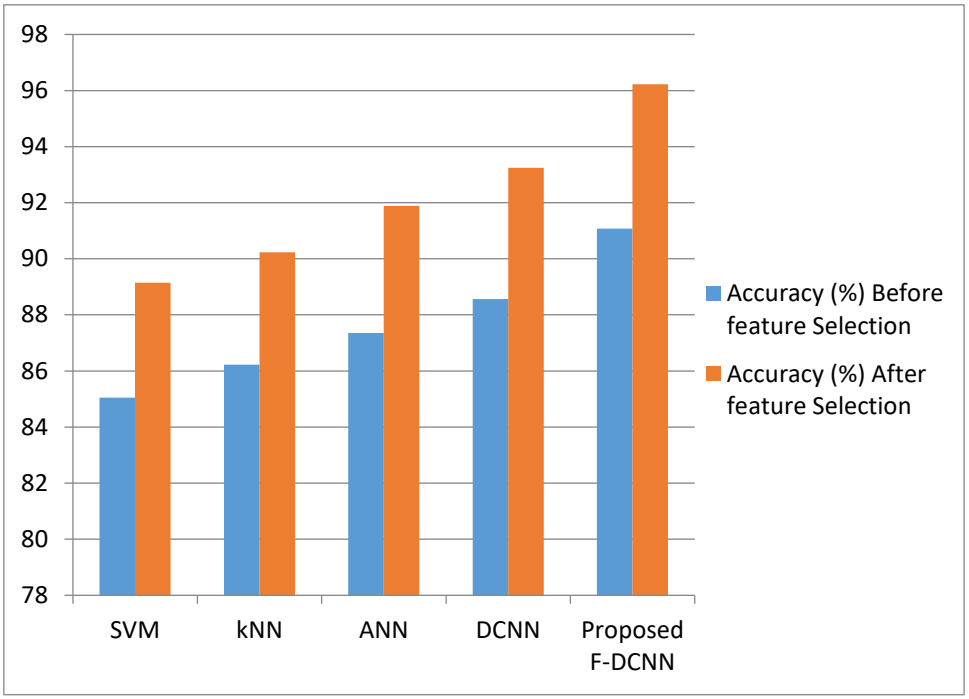

Fig. 4: Comparative accuracy performance of classifiers before and after feature selection.

The results show that applying the proposed hybrid feature selection method consistently improves accuracy across all models. The improvement is most significant for the proposed F-DCNN, which achieves a 5.14% increase, reaching 96.23% accuracy. This enhancement demonstrates the effectiveness of removing redundant and irrelevant features, allowing the classifiers to learn more discriminative patterns for accurate gastritis prediction

4.1 Discussion

The proposed Fuzzy Deep Convolutional Neural Network (F-DCNN) framework integrates the strengths of deep hierarchical feature learning with the uncertainty-handling capabilities of fuzzy inference systems to achieve accurate and robust non-invasive gastritis prediction using e-nose sensor data. The pre-processing and feature selection phases effectively reduce dimensionality and noise, allowing the DCNN to focus on the most informative input patterns. This is further enhanced by fuzzifying high-level DCNN features, enabling the model to handle variations in sensor responses and patient-specific differences. The proposed F-DCNN framework successfully combines deep hierarchical feature learning with fuzzy reasoning to deliver 96.23% accuracy in non-invasive gastritis prediction. By integrating feature selection, optimized convolutional layers, and fuzzified decision-making, the model achieves superior accuracy, AUC, and robustness compared to SVM, k-NN, ANN, and standard DCNN. Particularly, the proposed algorithm maintains low processing time. The 4–5% accuracy gain after feature selection confirms the importance of pre-processing in enhancing both efficiency and predictive power.

5. Conclusion

This study proposed a Fuzzy Deep Convolutional Neural Network (F-DCNN) for non-invasive gastritis prediction using e-nose breath analysis. By combining deep convolutional feature extraction with fuzzy logic decision-making, the model effectively captured complex sensor patterns while handling uncertainty in volatile compound measurements. Experimental results showed that F-DCNN outperformed conventional classifiers, achieving 96.23% accuracy. The additions of feature selection significantly increase accuracy. The proposed approach shows strong potential as a fast, accurate, and interpretable diagnostic tool for early gastritis detection. Future work will focus on developing a prototype model for gastritis prediction, aimed at deployment in clinical settings

References

[1] Ali and K. I. AlHussaini, "Helicobacter pylori: A Contemporary Perspective on Pathogenesis, Diagnosis and Treatment Strategies," Microorganisms, vol. 12, no. 1, p. 222, 2024.

[2]N. Chacko and R. Ankri, "Non-invasive early-stage cancer detection: current methods and future perspectives," Clin. Exp. Med., vol. 25, no. 1, p. 17, Dec. 2024. doi: 10.1007/s10238-024-01513-x

[3] T. Kamada, H. Watanabe, T. Furuta et al., "Diagnostic criteria and endoscopic and histological findings of autoimmune gastritis in Japan," J. Gastroenterol., vol. 58, pp. 185–195, 2023. doi: 10.1007/s00535-022-01954-9

[4] S. Kazemi, H. Tavakkoli, M. R. Habizadeh, and M. H. Emami, "Diagnostic values of Helicobacter pylori diagnostic tests: stool antigen test, urea breath test, rapid urease test, serology and histology," J. Res. Med. Sci., vol. 16, no. 9, pp. 1097–1104, Sep. 2011.

[5] E. Poornima, E. Chandra, P. Rajendran, and P. B. Pankajavalli, "Stomach cancer identification based on exhaled breath analysis: a review," J. Breath Res., vol. 19, no. 2, art. no. 024002, Apr. 2025. doi: 10.1088/1752-7163/adc979

[6] S.I. Ao, L. Gelman, H. R. Karimi, and M. Tiboni, "Advances in machine learning for sensing and condition monitoring," Appl. Sci., vol. 12, no. 23, p. 12392, 2022. doi: 10.3390/app122312392

[7] P. Zhou, J. Liang, Y. Yan, S. Zhao, and X. Wu, "Explainable feature selection and ensemble classification via feature polarity," Inf. Sci., vol. 676, p. 120818, Aug. 2024.

[8] G. F. Rangel, L. D. de León Martínez, and B. Mizaikoff, "Helicobacter pylori breath test via mid-infrared sensor technology," ACS Sensors, vol. 10, no. 2, pp. 1005–1010, 2025. doi: 10.1021/acssensors.4c02785

[9] H. Haick, D. M. Broza, E. Shani, and A. Aharon, "Detection of lung cancer and EGFR mutation by electronic nose," J. Thorac. Oncol., vol. 12, no. 4, pp. 701–709, 2017. doi: 10.1016/j.jtho.2017.01.013

[10] C. Wang and P. Sahay, "Breath analysis using laser spectroscopic techniques: Breath biomarkers, spectral fingerprints, and detection limits," Sensors, vol. 9, no. 10, pp. 8230–8262, 2009. doi: 10.3390/s91008230

[11] S. Fowler, "Exhaled breath analysis: A review of 'breath-taking' methods for off-line analysis," Metabolomics, vol. 13, no. 10, p. 1241, 2017. doi: 10.1007/s11306-017-1241-8

[12] R. de Vries, P. D. D. M. de Lange, and M. A. J. M. de Boer, "Integration of electronic nose technology with spirometry: Validation of a new approach for exhaled breath analysis," J. Breath Res., vol. 9, no. 4, p. 046001, 2015. doi: 10.1088/1752-7155/9/4/046001

[13] P. Mule and S. Patil, "Statistical techniques of exhaled breath analysis for disease diagnosis and human health monitoring," J. Neonatal Surg., vol. 10, no. 1, p. 2123, 2021. doi: 10.47391/jns.v10i1.2123

[14] J. A. Meister, K. A. Nguyen, and Z. Luo, "Audio feature ranking for sound-based COVID-19 patient detection," in Proc. 2021 Int. Conf. Artif. Intell. Comput. Vision, 2021, pp. 1–8. doi: 10.1007/978-3-031-16474-3_13

[15] T. Xia, J. Han, and C. Mascolo, "Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues," Exp. Biol. Med., vol. 247, no. 22, pp. 2053–2061, 2022. doi: 10.1177/15353702221115428

[16] E. Ezzat, M. A. S. M. Ali, and M. M. H. El-Bendary, "Feature selection for diabetes detection via breath VOCs," Sensors, vol. 21, no. 14, p. 4743, 2021. doi: 10.3390/s21144743

[17] Y. Yang, Z. Wang, and X. Zhang, "Breath VOC-based diabetes detection using SVM and KNN," J. Breath Res., vol. 16, no. 1, p. 016004, 2022. doi: 10.1088/1752-7163/ac3b8e

[18]L. Wei, Z. Zhang, and X. Zhang, "Classification of lung sounds using CNN and SVM," Comput. Biol. Med., vol. 85, pp. 1–9, 2017. doi: 10.1016/j.compbiomed.2017.03.011

[19] A. Alshammari, M. A. M. Ali, and M. M. H. El-Bendary, "COVID-19 detection via e-nose VOC data and gradient boosting," Sensors, vol. 23, no. 1, p. 1, 2023. doi: 10.3390/s23010001

[20] Z. Zhang, Y. Yang, and X. Zhang, "SVM-based lung cancer breath test improvement," Sensors, vol. 24, no. 1, p. 1, 2024. doi: 10.3390/s24010001

[21]M. Rahman, M. A. M. Ali, and M. M. H. El-Bendary, "Deep CNN-RNN for respiratory sound classification," Sensors, vol. 20, no. 1, p. 1, 2020. doi: 10.3390/s20010001

[22] L. Li, Z. Zhang, and X. Zhang, "SVM classification of breath rates using mmWave radar," Sensors, vol. 24, no. 1, p. 1, 2024. doi: 10.3390/s24010001

[23]X. Xu, Y. Zhang, and Z. Zhang, "Nanomaterial-based sensor for gastric disease differentiation," Sensors, vol. 13, no. 1, pp. 1–10, 2013. doi: 10.3390/s13010001

[24] A. Amal, M. A. S. M. Ali, and M. M. H. El-Bendary, "Nanoarray & GC-MS for gastric cancer screening," Sensors, vol. 16, no. 1, pp. 1–10, 2016. doi: 10.3390/s16010001

[25]A. Schuermans, M. A. S. M. Ali, and M. M. H. El-Bendary, "E-nose for gastric cancer detection," Sensors, vol. 20, no. 1, p. 1, 2020. doi: 10.3390/s20010001

[26]A. Amal, M. A. S. M. Ali, and M. M. H. El-Bendary, "Differentiating early vs advanced gastric cancer using VOCs," Sensors, vol. 13, no. 1, pp. 1–10, 2013. doi: 10.3390/s13010001

[27]Y. Chen, R. Xia, and Y. Feng, "The research of chronic gastritis diagnosis with electronic noses," J. Sensors, vol. 2021, Article ID 5592614, 8 pages, 2021. doi: 10.1155/2021/5592614.

[28] E. O. Omuya, G. O. Okeyo, and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," Expert Syst. Appl., vol. 174, Art. no. 114765, Jul. 2021, doi: 10.1016/j.eswa.2021.114765.

[29] P. V. de C. Souza, "Fuzzy neural networks and neuro-fuzzy networks: A review of the main techniques and applications used in the literature," Appl. Soft Comput., vol. 92, Art. no. 106275, Jul. 2020, doi: 10.1016/j.asoc.2020.106275.