

Enhancing Interpretability in Lung CT Image Clustering with HFMSTM-Enhanced Explainable CNN for Disease Detection

L. Dhanapriya¹, S. Preetha²

¹Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India
Email: dhanapriyaphd[at]gmail.com

²Associate Professor and Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India
Email: preethacs[at]srcw.ac.in

Abstract: *The importance of interpretability in medical image clustering has grown significantly as a foundation for trustworthy clinical decision-making and healthcare transparency in Artificial Intelligence systems. Although Convolutional Neural Networks (CNN) and other Deep Learning (DL) methods have demonstrated remarkable results in distinctive feature extraction from CT scans of the lungs, the majority of clustering algorithms are unreliable, leaving no information about the clustering process or the significance of the features. Furthermore, traditional clustering algorithms are unable to preserve the topologic relationships of complex image representations, thus making the clustering result difficult to interpret and unreliable. In order to address the above drawbacks, this study proposed a novel approach for lung Computed Tomography (CT) image interpretation and clustering called the HFMSTM Enhanced Self Organizing Map (SOM)–Explainable CNN (HSECNN) framework. To obtain informative deep feature representations, the proposed system employed an attention-based Explainable CNN, and to perform topology preserving clustering, an HFMSTM based SOM was used. The HFMSTM enhancement enabled neighborhood connectivity and cluster organization which enabled complex patterns in lung CT images to be well represented. In order to emphasize transparency even more, SHapley Additive exPlanations (SHAP) was employed for analyzing the global importance of features, while Local Interpretable Model-Agnostic Explanations (LIME) was used for providing local justifications for each cluster participation. Moreover, visualization modules were added to gain insight into clusters, feature distributions and topological relationships. The suggested HSECNN framework outperformed the other methods in terms of precision, recall, accuracy, and F1-Score, according to the experimental data. The performance of the clustering was improved through the use of topology-preserving clustering and Explainable AI (XAI) methods, which also helped to make the clustering transparent and trusted by users. As a result, medical image analysis and clustering of CT scans of the lungs are both improved by the suggested framework.*

Keywords: Lung CT Images, Interpretable Clustering, Explainable CNN, SOM, HFMSTM, SHAP, LIME, XAI.

1. Introduction

Interpretability has become increasingly important in medical image analysis since clustering outcomes directly influence clinical understanding and decision-making processes [1]. Although clustering techniques have demonstrated effectiveness in discovering hidden structures from lung CT images, understanding cluster assignments and identifying influential features remain challenging tasks [2,3]. Additionally, lung CT images have heterogeneous characteristics, complex texture patterns, and high-dimensional data, making the clustering of interpretable images challenging [4,5]. All of these restraints hinder visibility, validation, and user confidence in medical systems based on clustering.

The extraction of hierarchical representation of features from lung CT images has been a promising area of research for several DL approaches, particularly CNNs [6,7]. CNN architectures automatically learn discriminative spatial features and enhance clustering performance by representation learning [8]. However, despite all these benefits, CNN-based clustering systems often remain as black-box systems, limiting understanding of clustering behavior and reducing explainability of cluster assignments [9,10]. Consequently, insufficient interpretability restricts

their practical applicability in medical environments where transparent decision-making is essential [11,12].

SOM have found numerous applications to maintain topology and offer suitably low dimensional representations of high dimensional data [13,14]. SOM has significant advantages for better visualization of the clustering structure because it keeps the neighborhood relationships between the samples and helps preservation of topology in the mapping [15,16]. In the case of complicated medical imaging collections, however, conventional SOM approaches provide limited insight into the significance of cluster memberships and attributes. To improve transparency and increase user confidence, it is imperative to implement interpretable clustering frameworks that preserve topology while concurrently providing users with significant explanations of the clustering outcomes [11,17].

To overcome these limitations, this work proposes a novel HSECNN framework for interpretable clustering of lung CT images. The proposed framework combines HFMSTM [18] enhanced SOM with Explainable CNN architecture to enhance the clustering performance to ensure transparency. In addition, attention mechanisms are employed to improve the capability to represent features and discriminative learning. SHAP-based feature weighting is used to enhance interpretability by identifying features that are globally

significant in clustering decisions, whereas LIME-based explanations facilitate local interpretations of cluster memberships. Additionally, visualization modules are incorporated to improve understanding of cluster assignments, feature distributions, and intrinsic data topology. The proposed HSECNN system integrates clustering performance with explainability techniques to improve transparency, increases user trust, and facilitate interpretable medical image analysis.

Here is the structure of the remaining sections: Related research studies are summarised in Section 2. In Section 3, the suggested approach is laid out in detail. Section 4 evaluates the performance analysis and experimental outcomes. The study and its implications for future research are wrapped up in Section 5.

2. Literature Survey

Hasan et al., [19] developed a framework for interpreting lung ultrasound image using fuzzy pooling DL model for disease diagnosis. This approach used CNN architectures with fuzzy pooling layers to obtain discriminative features and then categorize lung ultrasound images into several disease categories. Additionally, SHAP-based explanations have been added to assist in the interpretation of prediction results through identification of feature importance in the intermediate layers of the network. Additional effort was necessary to implement the framework in more interpretable clinical decision-making applications, as it focused primarily on classification tasks, neglecting a thorough investigation of disease severity and an extensive topology-based interpretation.

Sudhish et al. [20] devised a content-based image extraction system for clinical diagnosis based on fuzzy grouping of DL. This model utilized CNN for obtaining features with a grouping method for categorizing the attribute mapping dataset. To remove the length of attribute vectors obtained through the previously trained CNN frameworks, a Multi-Level Gain-based choosing traits was used. This approach overlooked several intricacies in disease-specific patterns, which diminished accuracy in the context of uncommon disorders.

An improved method for forecasting and decision-making was proposed by Ahmed et al., [21] using a hybrid predictive model that integrates SOM, Principal Component Analysis (PCA), and Artificial Neural Networks (ANN) for better results. It adopted som as a feature extractor and as a topology preserving representation, PCA as a dimensionality reduction and ANN as a prediction and classification. Additionally, interpretability analysis (SHAP) was implemented to boost transparency in the decision-making process and shed light on the model prediction. The model was more interpretable and predictive than traditional machine learning models, leveraging structured historical data to achieve limited performance in real-time changes and unstructured data sources.

Rahman and Rahman [22] introduced an approach for demand forecasting that is based on XAI that uses along with ANN to

provide more accurate and interpretable predictions. The implemented cluster framework utilised SOM to discern clusters of diverse demand patterns, while NN was employed for non-linear forecasting inside these demand clusters. In addition, SHAP-based interpretability was applied to show the contribution of features at the local and global level when explaining the predictions of the model.

Danylenko & Yurchenko [23] devised an interpretable demand forecasting model for multi-echelon supply chains combining the use of a DL model, a SOM clustering approach with SHAP-based explainability. The framework used a SOM model to group the demand patterns and supply chain attributes which allowed to present a more complex structure of data before forecasting. Both LSTM and Transformer networks have been used to predict demand at various levels, and SHAP explanations are used to quantify the contribution of various factors, including lead time, seasonality and upstream demand, to the forecast results. The framework performed well in terms of interpretability and forecast accuracy, but it struggled to handle extremely unpredictable supply chain conditions since it failed to account for certain dynamic shocks and extreme demand changes.

Danylenko & Yurchenko [24] introduced a new stable demand forecasting system under uncertainty and introduced an explainable learning system in the forecasting process using ANN, XAI and SOM. It used SOM to group and represent complex demand patterns into clusters, and then reduced dimensionality and forecast the demands within each cluster. ANN was used to model nonlinear relationships and to make accurate demand prediction and XAI techniques were used to explain the importance of features, prediction confidence, and decision-making processes. The hybrid paradigm has helped enhance forecasting accuracy and greater transparency, thereby aiding in informed decision-making among sectors like retail, energy, and manufacturing. The framework did not adequately address some of the market disruptions that occurred over a short period or rare uncertainty patterns, which could have adversely impacted forecast accuracy during a volatile period.

3. Proposed Methodology

The section focuses on the development of the deep clustering model that is suggested to use HSECNN. Figure 1 depicts the pipeline of the proposed model.

3.1 Pre-processing

Improving visual quality and making sure the input to the suggested HSECNN framework is consistent are two major goals of data pre-processing. The images of the lungs from various sets may exhibit differences in the image resolution, intensity distribution, noise, and scanning protocol, potentially influencing the quality of feature extraction and clustering. A series of pre-processing operations is performed to improve image quality and to minimize data inconsistencies. The pre-processing procedure consists of the following stages:

- Noise Reduction: First, the noise reduction is used to

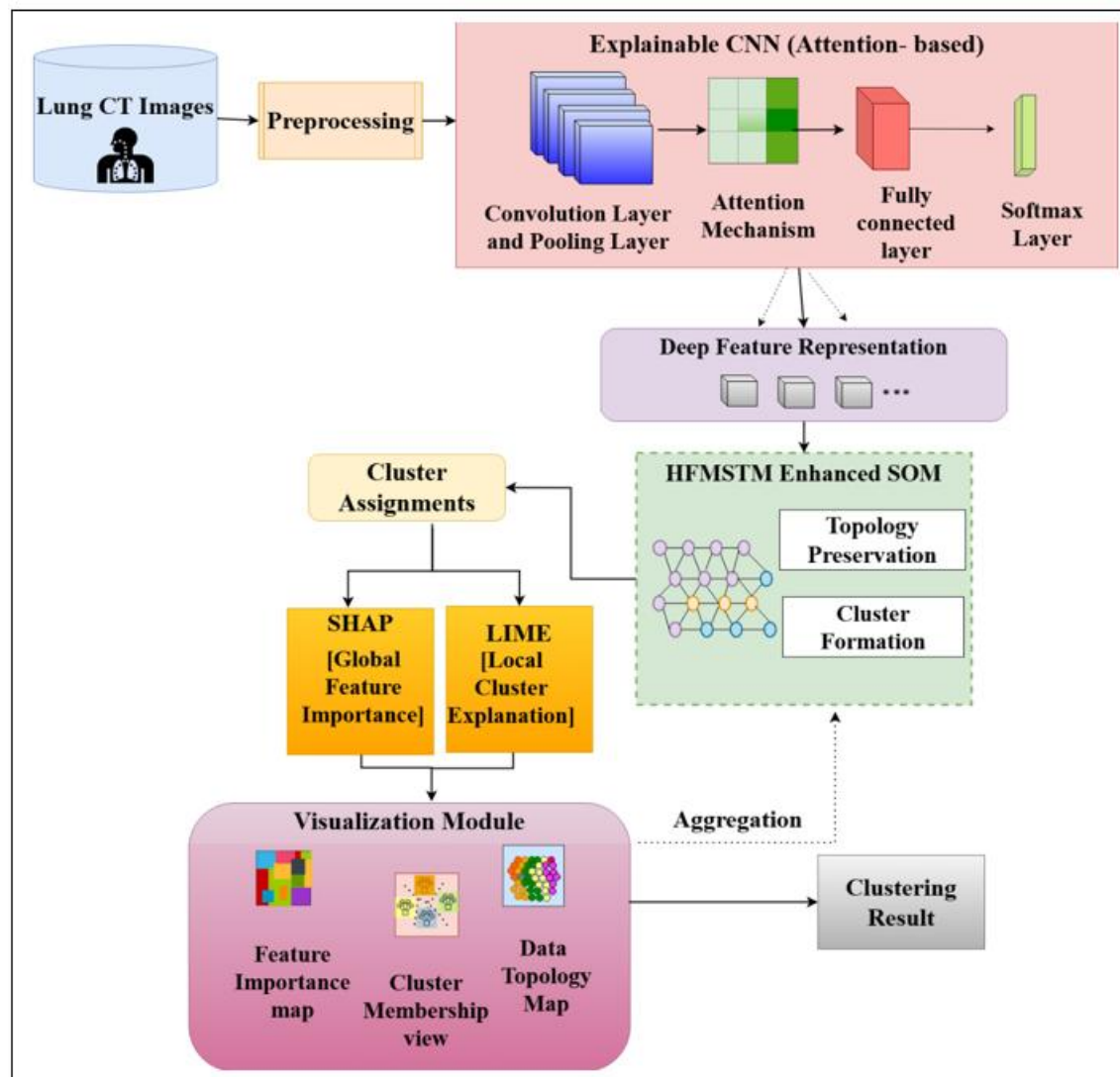


Figure 1: Pipeline of the proposed Model

minimize unwanted variations and image artifacts, but keep the important anatomical structures in the lung areas. This process will enhance the interpretability of the images and enable appropriate learning of features.

- Normalization: The denoised images are normalized to standardize the intensity values of the pixels and minimize variations resulting from the different CT acquisition settings. Normalization can make the images more consistent, thus making the learning process more robust.
- Resizing: To achieve the uniform input dimension and reduce the computational complexity, all CT images are resized to $224 \times 224 \times 3$ before feeding into the DL framework. This operation improves processing efficiency and enables consistent feature extraction across all samples.

These pre-processing steps enhance the image quality, reduce variations across different data sets and ensure it feeds into the Explainable CNN module in a pre-standardized manner. Therefore, extracted representations of features are more trustworthy and lead to more accurate clustering and interpretability results for the proposed framework HSECNN.

3.2 Explainable CNN for Deep Feature Extraction

The Explainable CNN (ECNN) is the feature extraction module of the proposed HSECNN framework. CNN architectures effectively learn hierarchical feature representations from medical images by automatically extracting discriminative visual characteristics at multiple abstraction levels. The proposed framework utilizes pre-processed lung CT image data, with dimensions of $224 \times 224 \times 3$, as input to the ECNN model. Layers such as fully connected, pooling, attention, convolutional and activation facilitate the extraction of discriminative spatial features and the production of relevant feature representations for use in later clustering operations. The first step in convolutional layers to find certain local spatial features in a CT image of the lungs, like edges and textures, as well as variations in intensity and disease-related structures. The convolution operation is mathematically expressed as Equation (1).

$$F_{i,j}^k = \sigma(\sum_m \sum_n W_{m,n}^k X_{i+m,j+n} + b^k) \quad (1)$$

When the input image is represented by X , the convolution kernel is W^k , the bias term is b^k , an activation function is σ , and the resulting feature map is $F_{i,j}^k$. The obtained feature maps capture meaningful visual features related to lung

abnormalities.

To further improve feature discrimination and introduce non-linearity, an activation function called a Rectified Linear Unit (ReLU) is employed. Equation (2) defines the ReLU operation.

$$ReLU(x) = \max(0, x) \quad (2)$$

where (x) represents the input activation value. Furthermore, an attention mechanism is built into the CNN architecture to selectively highlight the diagnostically important lung regions and to remove irrelevant background information, to further enrich feature learning. The attention map is calculated as Equation (3).

$$M(F) = \sigma(W_a F + b_a) \quad (3)$$

Here F is the feature map extracted, W_a is the weight matrix of attention, b_a is the bias term, and $M(F)$ is the attention map generated. It is possible to acquire the feature representation with attention enhancement via

$$F^* = M(F) \odot F \quad (4)$$

where F^* is the refined feature map and $\odot$ is element-wise multiplication. This attention mechanism highlights the diagnostically relevant part of the lungs and reduces the contribution of irrelevant background information, thereby allowing for the better extraction of the relevant lung disease features.

Later, the attention enhanced features are presented in a compact deep feature representation using fully connected layers. The deep features are a set of meaningful description of the CT images of lung and are used to identify complex image patterns. The proposed ECNN framework incorporates mechanisms for explaining CNN to make it more transparent and interpretable as opposed to conventional CNN models that are black-box systems. The resulting deep feature representations are then passed into the HFMSTM enhanced Self-organising Map module that performs topology-preserving clustering to cluster similar image patterns into meaningful clusters.

HFMSTM-Enhanced Self-Organizing Map for Interpretable Clustering

After deep feature extraction, the extracted features are passed to the HFMSTM-enhanced SOM module for clustering and topology preservation. Unsupervised SOM neural networks preserve neighbourhood associations between data samples while mapping high-dimensional feature vectors to low-dimensional maps. This topographic preservation property allows the complex pattern of medical images to be effectively visualised and understood. Unlike most of the clustering methods based on distance classification, SOM captures both local and global structural relationships, which improves cluster representation interpretability and quality.

The proposed HSECNN is designed to overcome the shortcomings of the conventional SOM framework by supplementing it with the HFMSTM mechanism, which provides better topology preservation, neighborhood connectivity, and cluster formation. The HFMSTM enhancement enhances neighborhood connectivity and retains intrinsic topological relationships between deep feature vectors, facilitating effective organization of complex patterns of lung CT images.

Picking a Best Matching Unit (BMU) with the smallest Euclidean distances between the input features vector and the weight vectors of the neurons is the first step. Equation (5) depicts the BMU selection procedure.

$$BMU = \underset{i}{\operatorname{argmin}} \|x - w_i\| \quad (5)$$

In this case, x is the input deep features vector and w_i is the weight vector for the i th neurone. Once the BMU is identified, a neighborhood function is used to update neighboring neurons, while retaining the topological relations between feature vectors. The neighborhood function is defined as

$$h_{ci}(t) = \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (6)$$

where r_c is the location of the winning neuron and r_i is the location of the neighboring neuron, respectively, while $\sigma(t)$ denotes the neighborhood radius at iteration (t) .

Subsequently, the neuron weights are adjusted to optimize cluster organization and topology preservation based on:

$$w_i(t+1) = w_i(t) + \alpha(t) h_{ci}(t) [x(t) - w_i(t)] \quad (7)$$

where the neighbourhood influence function is represented by $h_{ci}(t)$ and $\alpha(t)$ is the learning rate. With iterative adaptation, neighbouring neurons slowly learn to represent similar feature patterns and topology preserving cluster formation is gradually achieved. The architecture of the proposed HFMSTM-Enhanced SOM for interpretable clustering is shown in Figure 2.

The deep feature vectors are arranged with Hybrid Fuzzy Similarity Computation, Minimum Spanning Tree construction, and Manifold Clustering. The enhanced SOM performs BMU identification, neighborhood-based weight updating, and topology-preserving mapping to generate meaningful cluster formations and final cluster assignments.

The HFMSTM enhancement also improves SOM by enhancing neighborhood connectivity, maintaining the intrinsic topological relationships, and organizing the complex feature distribution. Lung CT images with similar pathological features are subsequently mapped to neighbouring neurons which help meaningful clustering and better topology visualization. The cluster assignment is then passed to the explainability module, where the SHAP and LIME modules are used to provide global feature importance and local explanations of the cluster membership, thereby enhancing transparency and interpretability in clustering outcomes.

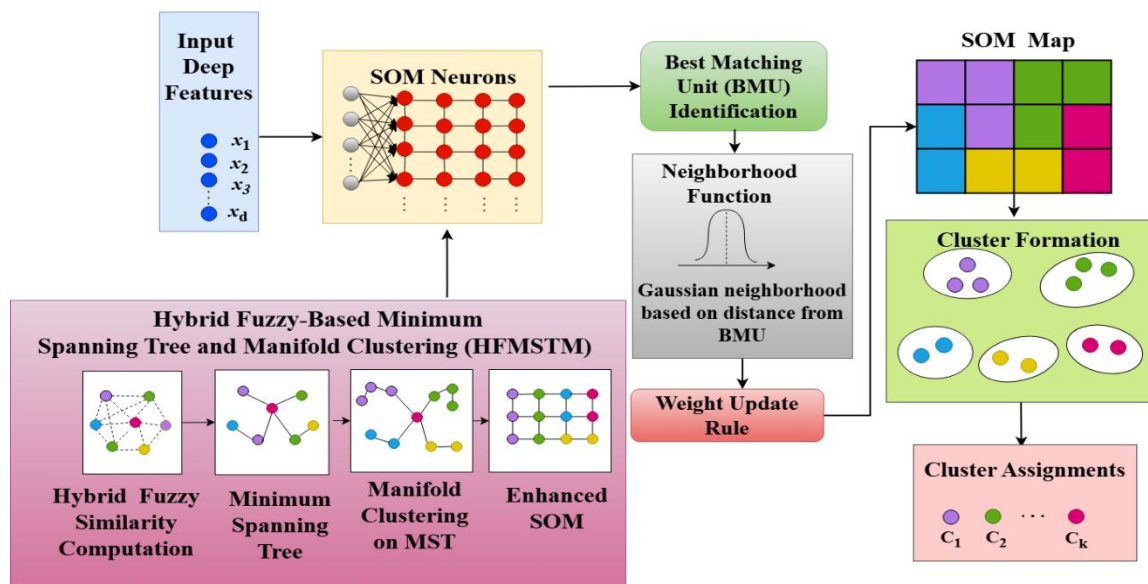


Figure 2: Structure of HFMSTM-Enhanced SOM

3.3 SHAP-Based Global Feature Importance Analysis

While the HFMSTM enhanced SOM successfully arranges the lung CT images in meaningful clusters, it is a difficult task to interpret the image features contributing to the decision of image cluster membership. To enhance transparency and interpretability, SHAP is integrated within the proposed HSECNN framework. SHAP is a game-theory-based explainability technique that quantifies the impact of all features on the final model output. To help comprehend the elements that affect clustering results, SHAP gives each characteristic an importance rating.

Following the proposed framework, SHAP is used to conduct global feature importance analysis of the deep feature representations extracted by the Explainable CNN and clustered by the HFMSTM enhanced SOM. Equation (8) is used to get the SHAP value of a feature:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (8)$$

A model's output function is denoted by $f(\cdot)$, where ϕ_i represents the SHAP value that represents feature i , N is the set of every features, and S is a subset of features that does not include feature i . The function that assesses the contribution of feature i to the clustering decision is denoted as $f(S \cup \{i\}) - f(S)$. By calculating the effect of each feature across all permutations, the SHAP value shows how much of a marginal contribution a feature makes.

Features are ranked by their effect on forming clusters based on the computed SHAP values. Features with high SHAP values have high influence on the outcome of the clustering process, while features with small SHAP have weak influence on the outcome. The global interpretation allows to detect the most significant deep features associated with lung CT image patterns, helping to understand the characteristics of the clusters.

Adding SHAP to HSECNN makes decision-making more transparent and provides quantitative insights into the

significance of features. Global explanations are generated that help the users understand the origins of cluster formation and boost their confidence in understanding the results of clustering. In the Visualisation and Interpretability Module, thorough interpretations of the clustering results are achieved by combining SHAP-based global explanations with LIME-based local explanations.

3.4 LIME-Based Local Cluster Membership Explanation

SHAP offers a global understanding of feature importance for the entire dataset, but cannot explain the reasons behind individual cluster assignments. To ensure instance-level interpretability, the suggested HSECNN employs LIME. As a post-hoc explainability method, LIME can generate local surrogate models that, for a given instance, imitate a complicated model. This not only facilitates the interpretation of the individual clustering decisions but also assists the viewer in understanding the reason each lung CT image has been categorized into a cluster.

In the proposed framework, LIME is applied to explain cluster memberships generated by the HFMSTM-enhanced SOM. LIME is designed to identify the most influential features contributing to the cluster assignment by perturbing the input features of a given lung CT image and analyzing the changes in these clustering outcomes. The LIME objective function is expressed as Equation (9).

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega_{(g)} \quad (9)$$

$L(f, g, \pi_x)$ measures the surrogate model's fidelity in corresponding to the original approach around the neighborhood of instance x , π_x stands for the locality factor, and $\Omega_{(g)}$ indicates the complexity reward applied on the surrogate model. Here, $\xi(x)$ remains the local corresponding explanation found at instance (x) , (f) serves as the original clustering method, and (g) is the accessible surrogate model chosen from the group of interpretable models (G) .

With the local explanations generated, LIME identifies the features that are contributing positively or negatively towards the cluster membership of a single lung CT image. These explanations enable users to understand the reasoning behind clustering decisions and provide detailed insights into the characteristics influencing a particular cluster assignment. Moreover, the local explanations and global feature importance information provided by SHAP can be combined to form a comprehensive interpretability framework.

The proposed model integrates LIME into the HSECNN framework, generating greater transparency at the individual sample level and additional possibilities for detailed analysis of cluster memberships. The LIME-based local explanations are then combined with the SHAP-based global feature importance analysis and incorporated into the visualization and interpretability Module to enable an interpretable analysis of clustering results. This enhances transparency, builds user trust and supports reliable lung CT image analysis.

3.5 Visualization and Interpretability Module

The interpretability of clustering results is essential for understanding the underlying patterns present in lung CT images and increasing user confidence in clustering decisions. The HFMSTM-enhanced SOM provides a meaningful grouping of similar samples, but it can still be difficult for the user to interpret cluster structures, contribution of features, and data topology. To overcome this drawback, a Visualization and Interpretability Module is added to the proposed HSECNN framework.

Data from the HFMSTM enhanced SOM, SHAP and LIME is merged in the visualization module to get complete insights into the result of the clustering. The HFMSTM-improved SOM generates topology preserving cluster maps, which aids visualization of relationships between the lung CT image samples. These maps can be used to highlight cluster boundaries, neighborhood structures and similarities among various pathological patterns.

In addition, SHAP visualisations like feature importance plots are used to detect the most influential deep features for the cluster formation and general clustering behaviour patterns. Users are better able to grasp the relevance of particular variables, and the way they affect clustering results with the help of these images.

The overall interpretability of clustering results is gained by combining SHAP-based global importance of features analysis with LIME-based explainability analysis of local clusters. The integrated explanations are then visualized to enhance the understanding of feature contributions, cluster memberships and topology preserving clustering structures.

Moreover, the explanations provided by LIME are visual and provide local explanations for cluster memberships. LIME detects characteristics that either positively or negatively impact the assignment of a user-selected lung CT scan. The instance-level explanation will give insight into the thinking behind particular clusters and confirm consistency of model decisions.

The topology preservation ability is used to further improve the visual analysis by the HFMSTM enhanced SOM to represent relations of the features in a low dimensional space. This enables users to investigate their intrinsic data structure, view distributions of clusters and discover similarities between lung CT image patterns. Both topology maps and SHAP explanations and LIME explanations provide a full picture of clustering results.

The proposed framework for HSECNN will use the visualization and interpretability Module to enhance transparency, model validation and user confidence in the clustering results. The generated visual explanations help researchers and healthcare professionals to understand cluster assignments, feature importance, and the topology of the data, which leads to reliable and interpretable lung CT image analysis.

Algorithm 1: HSECNN

Input: Lung CT image dataset

Output: Cluster assignments, feature importance information, local explanations, and visualization results for lung CT images

- 1) Begin
- 2) Acquire lung CT image datasets and pre-process the images.
- 3) Resize images to $224 \times 224 \times 3$ and normalize pixel intensity values.
- 4) Feed the pre-processed images into the Explainable CNN.
- 5) Apply convolution operations to extract local spatial features (Equation (1)).
- 6) Incorporate non-linearity using ReLU activation (Equation (2)).
- 7) Reduce the feature dimensions by doing max-pooling.
- 8) Improve the representations of discriminative features (Equations (3)(4)) by using the attention mechanism.
- 9) Generate deep feature vectors through the fully connected layer.
- 10) Initialize HFMSTM-enhanced SOM neuron weights.
- 11) Compute the Best Matching Unit (BMU) for each feature vector (Equations (5)).
- 12) Calculate the Gaussian neighborhood function for neighboring neurons (Equations (6)).
- 13) Update neuron weight vectors using the SOM learning rule (Equations (7)).
- 14) Enhance SOM topology preservation and neighborhood connectivity using the HFMSTM mechanism.
- 15) Organize similar lung CT image features into topology-preserving clusters.
- 16) Generate cluster assignment maps from the trained HFMSTM-enhanced SOM.
- 17) Compute SHAP values to determine global feature importance (Equations (8)).
- 18) Rank features according to their SHAP importance scores.
- 19) Apply LIME to generate local explanations for individual cluster memberships (Equations (9)).
- 20) Generate visualization outputs for topology preservation and cluster organization.
- 21) Determine the clustering performance by calculating the F1-Score, Accuracy, Precision, and Recall.
- 22) Produce meaningful cluster assignments and visualization.

23) End

4. Result and Discussion

This section assesses the suggested HSECNN framework by contrasting it with current models.

4.1 Dataset Description

CT Scans for a Lung Cancer Database [25]: For the purpose of lung cancer diagnosis and classification, there is a publicly available dataset called CT Scan Image for Lung Cancer. This dataset contains CT scan images of persons with different lung issues. In this research, 100 images of Benign, 561 images of Malignant and 631 images of Normal were selected from 1292 images for pulmonary disease classification and clustering.

Customized CT Image Dataset:

For experimental analysis, lung CT images were collected from multiple publicly available datasets, including the MIDRC-RICORD-1a dataset [26], the SIIM-ACR Pneumothorax Segmentation dataset [27], the MultiClass Pulmonary Disease CT Image Dataset [28], and the Lungs Disease Dataset [29]. Lung CT images corresponding to six disease classes, namely Atelectasis, Edema, Normal, Pneumothorax, Pneumonia, and Tuberculosis, were collected from these datasets. A total of 2347 lung CT images

corresponding to six disease classes were selected and used for experimental evaluation.

Due to the extreme imbalance in both datasets, data augmentation techniques such as flipping, scaling, brightness adjustment, and rotation were employed to strengthen minority classes and enhance the model's generalizability. Following enhancement, the Lung Cancer CT now contains 6234 images, broken down as follows: 1989 images of benign, 2118 images of malignant, and 2127 images of normal. Likewise, the MultiClass Pulmonary Disease CT data set expanded to 3508 images of normal (603 images), atelectasis (541 images), edema (586 images), pneumothorax (582 images), pneumonia (649 images), and tuberculosis (547 images). Each augmented dataset contained 80% training samples and 20% testing samples for training and evaluation. The distribution of datasets class-wise before as well as following augmentation is shown in Tables 1 and 2, respectively.

4.2 Experimental Setup and Parameters Setting

A machine running Windows 10, equipped with the Intel Core TM i7-8700 processor, 16GB of RAM, and an NVIDIA GTX-1060 GPU was used to conduct the experiment. The programming platform is Python 3.12 and the DL framework is Tensorflow 1.6. The parameters of HSECNN model are enumerated in Table 3.

Table 1: Distribution of Lung Cancer CT Image Dataset

Categories	Pre augmentation			Post augmentation		
	Training	Testing	Total	Training	Testing	Total
Benign	80	20	100	1591	398	1,989
Malignant	449	112	561	1694	424	2,118
Normal	505	126	631	1702	425	2,127
Total	1034	258	1292	4987	1247	6234

Table 2: Distribution of Customized CT Image Dataset

Categories	Pre augmentation			Post augmentation		
	Training	Testing	Total	Training	Testing	Total
Atelectasis	75	19	94	433	108	541
Edema	61	15	76	469	117	586
Normal	328	82	410	482	121	603
Pneumothorax	431	108	539	466	116	582
Pneumonia	653	163	816	519	130	649
Tuberculosis	330	82	412	438	109	547
Total	1878	469	2347	2807	701	3508

Table 3: Simulation parameters

Model	Hyperparameter	Settings
Explainable CNN	Learning Rate	0.001
	Dimension of Kernel	3×3
	Pooling	2×2
	Size of batch	32
	Optimizer	Adam
	Activation	ReLU
	Epochs	100
HFMSTM-Enhanced SOM	Grid Size	10×10
	Learning Rate	0.5
	Iterations	1000
	Neighborhood Function	Gaussian
	Initial Neighborhood Radius	5
SHAP	Explanation Method	Global Feature Importance
LIME	Kernel Width	0.75

Volume 15 Issue 7, July 2026

Fully Refereed | Open Access | Double Blind Peer Reviewed Journal

www.ijsr.net

4.3 Evaluation Measures

The True Negatives are denoted as TN, the False Negatives as FN, the True Positives as TP, and the False Positives as FP.

- **Accuracy:** The percentage of samples that were accurately predicted relative to the overall number of samples that were analyzed. The degree to which the model was successful in generating accurate results is called its accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

- **Precision:** The accuracy rate is the proportion of the model's predictions that turn out to be correct, in this case, samples of lung illnesses.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

- **Recall:** It finds the fraction of positive cases in the dataset that were really predicted to be positive.

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

- **F1-Score:** It strikes a balance between recall and precision

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

4.4 Analysis of Performance

Existing models such as FP-CNN [18], FC-CNN-CBIR [19], and SOM-LSTM-SHAP [22] are compared with the proposed HSECNN in this section. Using the effectiveness criteria outlined in section 4.3, these models are tested on the dataset provided in section 4.1.

4.4.1 Evaluating the Performance of CT Scan Images Utilising the Lung Cancer Database

With a three-class classification dataset consisting of CT scans images of lung cancer, the suggested model achieved outstanding classification performance, as shown in Figure 3. The majority of samples fall along the diagonal, corresponding to correct predictions, and a few samples were misclassified. Figure 4 shows the learning efficiency of the suggested HSECNN model over 25 epochs, using accuracy in training and testing for CT scans of the lung with cancer. Excellent learning and convergence are demonstrated by the fact that both the testing and training accuracy consistently improve with each epoch. The training accuracy is raised from around 69% to 97.5%, and the testing accuracy is raised from 65% to 96%. The small difference between training and testing accuracy at the last epoch shows that the proposed HSECNN architecture ensures good and reliable clustering results on lung CT images without losing generalization to unseen data.

With regard to the precision, recall, accuracy, and F1-score criteria, Figure 5 compares the effectiveness of FP-CNN, FC-CNN-CBIR, SOM-LSTM-SHAP, and the proposed HSECNN. The alternative CNN models, FP-CNN, FC-CNN-CBIR, and SOM-LSTM-SHAP, all have lower accuracy rates than the suggested HSECNN (97.75 % vs. 9.60 %, 7.66 %, and 4.14%, respectively). Likewise, HSECNN achieves Precision of 97.74%, which is an improvement of almost 9.72%, 7.83%, and 4.28% over the existing models. In Recall, the proposed model records accuracy of 97.76%, which is higher by approximately 9.67%, 7.75%, and 4.18% in comparison with FP-CNN, FC-CNN-CBIR, and SOM-LSTM-SHAP, respectively. Moreover, the F1-score of the proposed F1CNN is 97.75%, which shows improvement of about 9.57%, 7.79%, and 4.33% over FP-CNN, FC-CNN-CBIR, and SOM-LSTM-SHAP, respectively.

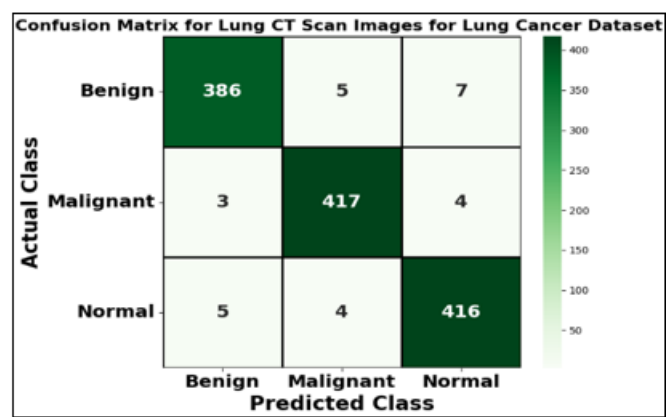


Figure 3: A Confusion matrix representing the CT Scan Images used in the Lung Cancer Dataset and the Suggested model

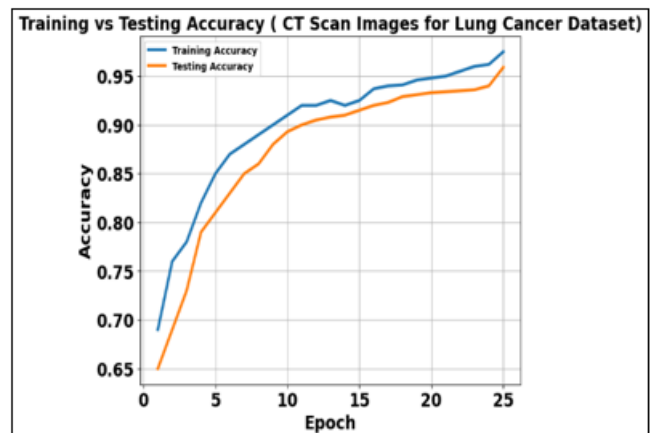


Figure 4: Evaluation of the suggested model's accuracy on a training set of CT Scan Images from a Lung Cancer Dataset compared to a testing set

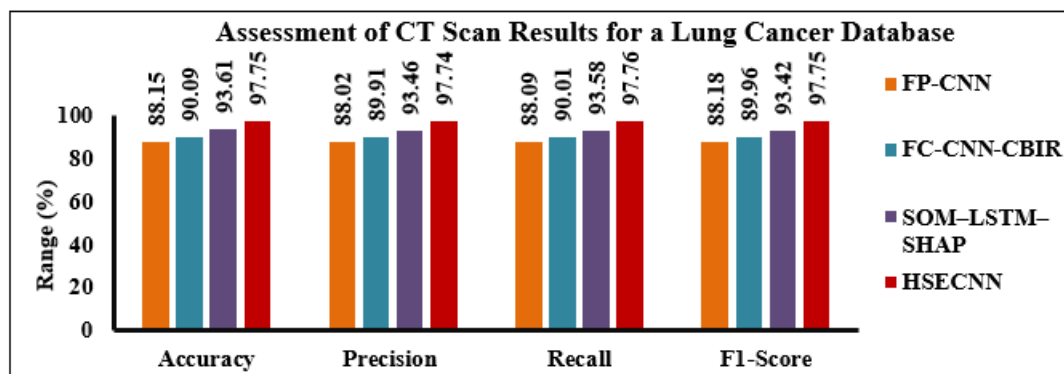


Figure 5: Classification study of current models and the proposed HSECNN using CT Scan Images from the Lung Cancer Dataset

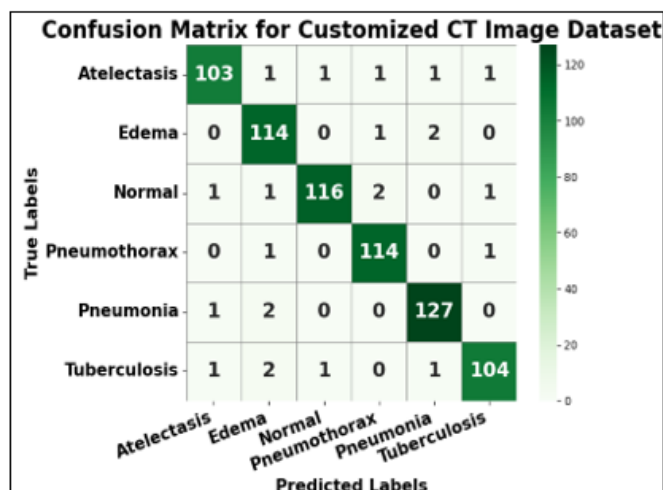


Figure 6: A Confusion matrix representing the CT Scan Images used in the Customized CT Image Dataset and the suggested model

4.4.2 Performance Evaluation on Customized CT Image Dataset

The suggested model's confusion matrix based on the Customized CT Image Dataset (6 distinct classes) appears in Figure 6. For most of the samples, the values are clustered around the diagonal, reflecting very accurate classification.

Figure 7 illustrates the Training vs Testing Accuracy curves of the proposed HSECNN model for the Customized CT

Image dataset for 25 epochs. The training and test accuracy improve over epochs, suggesting good learning and stable convergence. There is an increase of 62% to 95% in testing accuracy and an improvement of about 65% to 96.7% in training accuracy. Good generalizability with little overfitting is seen by a small difference between the two curves. There was a strong and consistent enhancement of clustering ability on the lung CT images, and the suggested HSECNN framework was able to transfer this improvement to the unknown dataset with minimal training and testing mistakes at the last epochs.

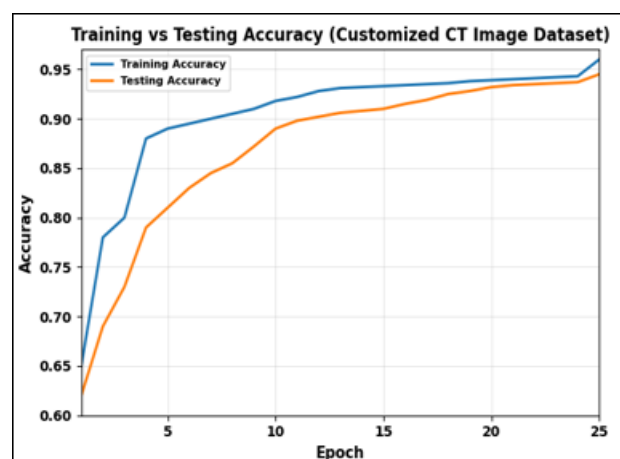


Figure 7: Performance comparison for training and testing set on Customized CT Image Dataset in terms of accuracy using proposed model

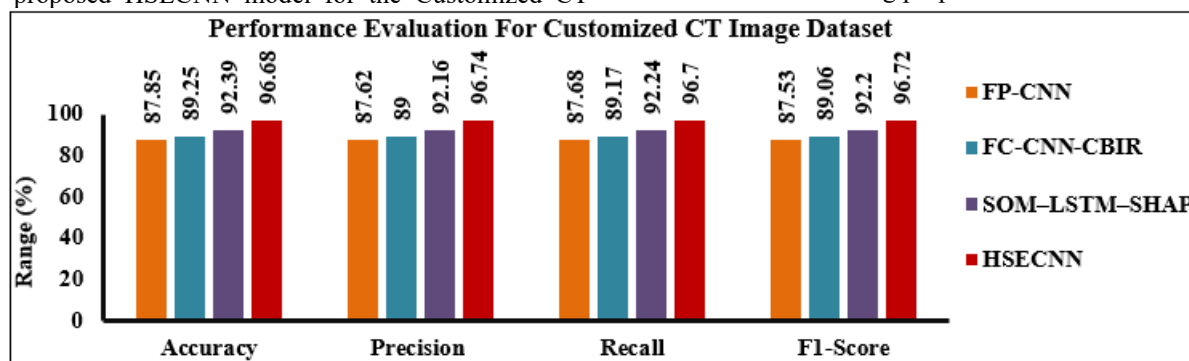


Figure 8: Classification study of current models and the proposed HSECNN using Customized CT Image Dataset

Considering several performance evaluation criteria like Accuracy, Precision, Recall, and F1-score, Figure 8 compares the performance of FP-CNN, FC-CNN-CBIR, SOM-LSTM-SHAP, and the proposed HSECNN. The suggested HSECNN

outperforms the other models by a significant margin: 8.83% over FP-CNN, 7.43% versus FC-CNN-CBIR, and 4.29% versus SOM-LSTM-SHAP. Likewise, the Precision of HSECNN is improved to 96.74% which shows the increase of

4.58%, 7.74%, and nearly 9.12% from the existing models. In the case of Recall the proposed model yields 96.70%, which outperforms FP-CNN, FC-CNN-CBIR, and SOM-LSTM-SHAP by 9.02%, 7.53%, and 4.46%, respectively. Moreover, performance improvements of around approximately 9.19%, 7.66%, and 4.52% are obtained with the F1 score over FP-CNN, FC-CNN-CBIR, and SOM-LSTM-SHAP, respectively.

5. Conclusion

In this study, a novel HSECNN framework for interpretable clustering of lung CT images was proposed. This framework consisted of an attention-based Explainable CNN module that learned discriminative feature deep features and the topology-preserving clustering was obtained by an HFMSTM, which was enhanced with SOM. The proposed method is to organize the complex patterns of lung CT images in neighbourhoods and represent the features in the neighbourhood relationships. To make the clustering findings more interpretable and the model more transparent, SHAP and then LIME were additionally implemented. Global importance of features analysis was performed using SHAP, while local descriptions of cluster memberships were performed using LIME. In comparison to the state-of-the-art approaches, the suggested HSECNN framework achieved significantly higher precision, recall, accuracy, and F1-Score in the experiments. The quality of the cluster and topology preservation was improved from the HFMSTM enhancement, while the feature representation learning was improved by the attention mechanism. Furthermore, visualization modules helped understanding the structures of clusters, the distributions of features and clustering decisions. The outcomes demonstrated that the suggested framework for analyzing lung CT images is effective, robust, and interpretable. Overall, the HSECNN approach successfully addressed the limitations of existing clustering methods, by integrating topology-preserving clustering and explainable AI techniques. In future, the multimodal medical data could be incorporated into the system, and attention mechanism could be further optimized to provide higher accuracy for clustering, scalability, and clinical applicability.

References

- [1] A. Singh, S. Sengupta, V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of imaging*, VI (6), p. 52, 2020.
- [2] K. Huang, L. Yang, Y. Wang, L. Huang, X. Zhou, W. Zhang, "Identification of non-small-cell lung cancer subtypes by unsupervised clustering of CT image features with distinct prognoses and gene pathway activities," *Biomedical Signal Processing and Control*, LXXVI, p. 103643, 2022.
- [3] K. Abdullahi, K. Ramakrishnan, A. B. Ali, "Deep learning techniques for lung cancer diagnosis with computed tomography imaging: A systematic review for detection, segmentation, and classification," *Information*, XVI (6), p. 451, 2025.
- [4] K. Wu, P. Wu, K. Yang, Z. Li, S. Kong, L. Yu, S. Wu, "A comprehensive texture feature analysis framework of renal cell carcinoma: pathological, prognostic, and genomic evaluation based on CT images," *European Radiology*, XXXII (4), pp. 2255-2265, 2022.
- [5] Z. Li, M. Zhao, Z. Li, Y.H. Huang, Z. Chen, Y. Pu, G. Ren, "Quantitative texture analysis using machine learning for predicting interpretable pulmonary perfusion from non-contrast computed tomography in pulmonary embolism patients," *Respiratory research*, XXV (1), P. 389, 2024.
- [6] Y. Wang, H. Zhang, K.J. Chae, Y. Choi, G.Y. Jin, S.B. Ko, "Novel convolutional neural network architecture for improved pulmonary nodule classification on computed tomography," *Multidimensional Systems and Signal Processing*, XXXI (3), PP. 1163-1183, 2020.
- [7] M.A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, S.R. Abdani, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images," *Diagnostics*, XIII (16), P. 2617, 2023.
- [8] C.C. Hsu, C.W. Lin, "Cnn-based joint clustering and representation learning with feature drift compensation for large-scale image data," *IEEE Transactions on Multimedia*, XX (2), PP. 421-429, 2017.
- [9] R. Azim, N. Abbas, H.K. Alkahtani, A. Qahmash, "An explainable deep learning framework for video violence detection using unsupervised keyframe selection and attention-based CNN," *Scientific Reports*, 2026.
- [10] S. Iftikhar, N. Anjum, A.B. Siddiqui, M. Ur Rehman, N. Ramzan, "Explainable CNN for brain tumor detection and classification through XAI based key features identification," *Brain Informatics*, XII (1), P. 10, 2025.
- [11] A. Chinnaraju, "Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability," *World Journal of Advanced Engineering Technology and Sciences*, XIV (3), PP. 170-207, 2025.
- [12] Q. Xu, W. Xie, B. Liao, C. Hu, L. Qin, Z. Yang, A. Luo, "Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: a systematic review," *Journal of healthcare engineering*, MMXXIII (1), P. 9919269, 2023.
- [13] N. Palli, "Enhancing Neural Network-based Dimensionality Reduction: A Focus on Self-Organizing Maps (SOM)," 2025.
- [14] A. Saraswati, M. Hagenbuchner, A.C. Tsoi, "High-resolution Self-Organizing Maps for advanced visualization and dimension reduction", *Neural Networks*, CV, PP.166-184, 2018.
- [15] K. Tasdemir, E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *IEEE Transactions on Neural Networks*, XX (4), PP. 549-562, 2009.
- [16] J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, J. Garcia-Rodriguez, H. Mora-Mora, "Constrained self-organizing feature map to preserve feature extraction topology," *Neural Computing and Applications*, XXVIII (Suppl 1), PP. 439-459, 2017.
- [17] G. Chen, J. Yuan, Y. Zhang, H. Zhu, R. Huang, F. Wang, W. Li, "Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery," *IEEE access*, XII, PP. 103348-103379, 2024.
- [18] L. Dhanapriya, S. Preetha, "HFMSTMC: Improved Binary Grey Wolf Optimization in Hybrid Fuzzy

- Minimum Spanning Tree Clustering with Manifold Learning,” ITEGAM-JETIA, XII (58), PP. 131-138, 2026.
- [19] M.M. Hasan, M. M. Hossain, M. M. Rahman, A.K.M. Azad, S.A. Alyami, M.A. Moni, “FP-CNN: Fuzzy pooling-based convolutional neural network for lung ultrasound image classification with explainable AI,” Computers in Biology and Medicine, CLXV, P. 107407, 2023.
- [20] D.K. Sudhish, L.R. Nair, “Content-based image retrieval for medical diagnosis using fuzzy clustering and deep learning,” Biomedical Signal Processing and Control, LXXXVIII, P. 105620, 2024.
- [21] K.R. Ahmed, M.E. Ansari, M. N. Ahsan, A. Rohan, M.B. Uddin, M.A.H. Rivin, “Deep learning framework for interpretable supply chain forecasting using SOM ANN and SHAP: KR Ahmed et al,” Scientific Reports, XIV (1), P. 26355, 2025.
- [22] F. Rahman, O.K.O.D. S. Rahman, “Explainable Artificial Intelligence for Supply Chain Forecasting: A SHAP-Driven Evaluation of SOM-ANN Frameworks”, 2025.
- [23] M. Danylenko, O.K.O.D.T. Yurchenko, “Interpretable Deep Learning for Multi-Echelon Supply Chain Forecasting Using SOM Clustering and SHAP Explanations”, 2025.
- [24] M. Danylenko, O.K.O.I. Holub, “Integrating SOM and ANN for Temporal Pattern Recognition in Supply Chain Forecasting: An Explainable AI Approach”, 2025.
- [25] R. Dishan, " CT Scan Images for Lung Cancer,". [Online]. Available: <https://www.kaggle.com/datasets/dishantrathi20/ct-scan-images-for-lung-cancer>.
- [26] Customized Image Dataset, Available: <https://www.cancerimagingarchive.net/collection/midrc-ricord-1a/>
- [27] <https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation>
- [28] <https://www.kaggle.com/datasets/programmer3/chest-diseases-by-medical-imaging>
- [29] <https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types>