# Fuzzy Consensus Clustering for Deep Learning Tuning. Taking Breast Cancer for Medical Diagnosis as a case

Choukri Djellali[1][0000−0003−2051−7894], Mehdi Adda[2][0000−0002−5327−1758], and Mohamed Tarik Moutacalli[2][0000−0003−3534−8624]

[1] Department of Computer Science
University of Quebec At Montreal
405 Rue Sainte-Catherine Est, QC H2L 2C4
Montréal, Canada,
djellali.choukri@courrier.uqam.ca
https://uqam.ca/
[2] Department of Mathematics, Computer Science and Engineering
University of Quebec At Rimouski
300 Allée des Ursulines, QC G5L 3A1
Rimouski, Canada
{mehdi_adda,mohamedtarik_moutacalli}@uqar.ca
https://www.uqar.ca/

**Abstract.** Deep Learning is a new branch of Machine Learning that focuses on applying Artificial Neural Networks to obtain more cluttered decision boundaries. However, the supervised Neural Networks are sensitive to presentation order, architecture configuration, and complex shapes. In addition, the learning instability causes large changes in its performance on training samples. In our study, we proposed a new model for RBF Neural Network tuning by using fuzzy consensus clustering and model selection. Experimental studies showed that our Deep Learning model, which is named DeepRBF, yielded better recognition accuracy than other supervised learning algorithms. In addition, our typical initialization scheme speeds up the learning convergence and avoids the local minimum.

**Keywords:** Medical Diagnosis · Data Mining · Deep Learning · Categorization · Artificial Neural Networks · Fuzzy logic · Model selection.

## 1 Introduction

Deep Learning (also known as Deep Neural Learning or Deep Neural Network) has recently attracted considerable attention due to its ability to learn complex decision boundaries by using Artificial Neural Networks (or ANNs). It is a Data Mining technique that trains Artificial Neural Networks by learning samples. Deep Learning has been used by several engineering systems, including Medical diagnosis [2], [19], Information Retrieval [13], Robotic [14], E-learning [11], Finance [17], Semantic Web [1], Digital telecommunication [22], and Autonomous

system [9], etc. All these disciplines show the practical importance of Deep Learning models.
Radial Basis Function Network (or RBF) [16] is a well-known Artificial Neural Network used in pattern categorization, and it has been extensively applied in Deep Learning. However, this supervised Neural Network is sensitive to architecture configuration, Bellman's curse of dimensionality, presentation order, and complex shapes. In addition, one of the biggest challenges of RBF learning is the learning instability.

On one hand, consensus clustering (also known as clustering ensemble or clustering aggregation) is a common technique for Data Mining, which has been developed with the objective of selecting a consolidated clustering model that is a better fitting of data.

On the other hand, model selection is a well-known data-based technique used in Data Mining to produce a stable model with the best inductive bias.

Based on these premises, we introduced a new scalable scheme for RBF tuning using an efficient fuzzy consensus clustering and model selection.

This paper is divided into six sections, including the introduction. In Section 2, we give an overview of the state of the art in this area. We discuss our research questions and the drawbacks of current categorization models. The conceptual architecture of our approach is described in detail in Section 3. Before we conclude, we give in Section 4 a short evaluation with benchmarking models for our conceptual model. Then, a conclusion (Section 5) ends the paper with future work (Section 6).

The next section investigates some of the more commonly used Deep Learning architectures and explores the rules of supervised learning for pattern categorization.

## 2 State of the art, problem and open issues

Deep Learning has become a useful technique with great potential to learn complex decision boundaries by using Artificial Neural Networks. The neural architecture contains a large number of interconnected artificial neurons, which are designed to mimic the units of the nervous system. Each artificial neuron computes a non-linear function of the inputs multiplied by the corresponding weights.
The Neural Network architecture is controlled by a Machine Learning mechanism. In supervised learning, the labels are already known, and the goal of learning is to infer a prediction function, which is used to return the predicted class labels for new samples. However, unsupervised learning looks for correlations between patterns to describe the relevant model without prior training.
There are a large number of Artificial Neural Networks that can be used for pattern categorization, particularly those used in Deep Learning. Among them, there is a class of Recurrent ANN Models [15] (e.g. Hopfield, Long Short- Term Memory (or LSTM), Bidirectional Associative Memory (or BAM), Deep Bidi-

rectional RNNs, Elman, etc.), Multilayer Feed-Forward ANN Models [24] (e.g. Backpropagation Through Time (or BPTT), Non-linear Auto-Regressive with eXogeneous inputs (or NARX), Multi- Layer Perceptron (or MLP), etc.), Convolutional Network or Space Invariant Networks [23] (e.g. LeNet, Residual Network (or ResNet), AlexNet, VGGNet, Convolutional Restricted Boltzmann Machines (or CRBMs), Convolutional Deep Belief Networks (or CDBNs), etc.), Recursive Neural Networks [12] (or RecNN) (e.g. Recursive auto-associative memories, Max-Margin NN, etc. ), etc.
This advancement has led to the appearance of several approaches for pattern classification.

Dsouza et al. [6] (2022) proposed a new hybrid Deep Learning model based on deep CNN (Convolution Neural Networks) architectures. The overfitting is avoided by using transfer techniques. This model used TensorFlow-GPU framework to accelerate the CNN training. The architecture configuration contained the parallel Short-Term Memory (LSTM) layers with MobileNet, VGG16, and ResNet. The evaluation, based on Histopathology image classification, showed that the Hybrid ResNet-based architectures with 101 and 152 yielded good results.

Ekman et al. [7](2025) presented a Deep Learning framework for 3D medical segmentation, by using three 2D U-Nets. The final solution is calculated by averaging the results from 2D U-Nets. The data used in this study comes from either CT or MR scans. Experimental studies are based on six clinical applications: orthopedics, orbital segmentation, mandible CT, cardiac CT, fetal MRI, and lung CT. Hyperparameter tuning and augmentation settings achieved an enhanced segmentation with an average Dice score of 92%, SD=$\pm$ 0.06.

Dang et al. [4] (2024) introduced an Ensemble scheme based on two-layer Deep Learning models for medical image segmentation. The layered architecture is used to find the optimal decision boundaries for model segmentation. The optimal configuration of this scheme facilitates the successive refinements of segmentation. The evaluation, based on five benchmark data sets, showed that the proposed model performed better than well-known benchmark algorithms.

Chen et al. [3] (2025) proposed a breast cancer classification based on multimodal medical images. This model used two channels of ResNet-18 for feature extraction. A fully connected neural network maps the extracted features. Sensitivity, Specificity, Accuracy, and Precision metrics assess model performance. The evaluation showed that the proposed model performed better than ResNet-18, ResNet-50, ResNext-50, Inception v3, VGG16, and GoogleNet.

A recent work [20] (2024) introduced a Deep Learning model for lung cancer detection. To enhance the contrast of Medical images and improve the classification accuracy, this model used histogram equalization. Threshold Segmentation is performed to reduce space representation. ReduceLROnPlateau, ModelCheckpoint, and EarlyStopping are used to adjust the learning rate, thereby optimizing the training process and improving the generalization.

Most categorization models are perturbed by presentation order, architecture configuration, and noisy patterns. Moreover, there are usually several ways

of pattern categorization, and the decision boundaries generated are not well separated.

In this paper, we describe a new Deep-learning-based categorization model for RBF Neural Network tuning, by using supervised Artificial Neural Networks, fuzzy consensus clustering, and model selection.
The next section explores how patterns are moved across our Deep Learning model and introduces the rules of supervised learning, consensus clustering, and model selection.

## 3 Architecture of our Deep Learning model

As shown in Figure 1, our Deep Learning model starts by presenting all patterns from a training set. Because the noise typically generates errors in the decision boundaries by defining nonexistent correlations between patterns, pretreatment is an extremely useful and time-saving task for Deep Learning.

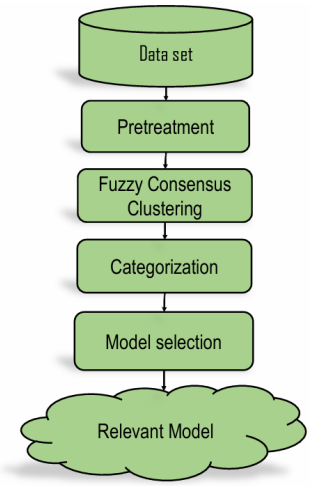


**Fig. 1.** The conceptual model.

We used Radial Basis Function Neural Network to categorize patterns according to their contents. It is a connectionist classification model that maps input patterns onto a set of predictors. This feed-forward neural architecture consists of multiple layers of neurons, which are fully connected. The RBF architecture depicted in Figure 2 consists of neurons arranged in three layers, namely, the input, hidden (RBF), and output layers.

Each neuron in the input (or hidden) layer is connected to all neurons in the hidden (or output) layer, respectively.

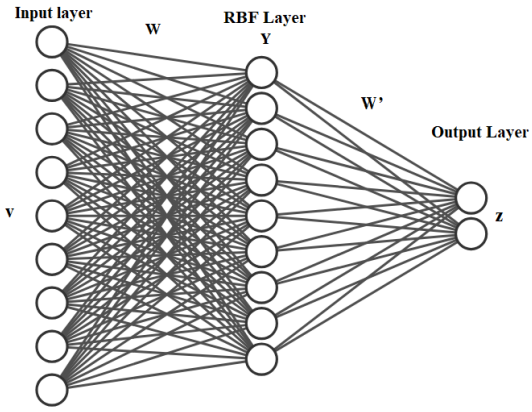


**Fig. 2.** RBF Architecture.

Each neuron in a hidden layer receives the outputs of all neurons in the input layer. The $i^{th}$ neuron activation is defined as follow:

$$a_i = \sum_{j=1}^{n} v_j w_{ji} \qquad y_i = f(a_i) = \frac{\|a_i - \mu_i\|}{2\sigma_i^2} \tag{1}$$

Where,
$\mu_i$ and $\sigma_i^2$: are respectively the center and variance of the Gaussian function.
The $i^{th}$ standard deviation is calculated as the average distance between patterns $p_i \in C_j$, and the cluster center $\mu_i$, i.e.,

$$\sigma_i = \frac{1}{\mid C_j \mid} \sum_{p_i \in C_j} \|p_i - \mu_i\| \tag{2}$$

The output neurons implement weighted sums of hidden neurons outputs, i.e.,

$$z_i = \sum_{j=1}^{p} y_j w_{ji}^{'} \tag{3}$$

The Mean Squared Deviation (or MSD) algorithm is used to update the synaptic weights, as defined by the following formula:

$$w_{ij}^{'(t+1)} \leftarrow w_{ij}^{'(t)} + \epsilon \sum_{j=1}^{m} (z_i^{desired} - z_i^{real})^2 \tag{4}$$

Where,
$\epsilon$: learning rate.
$z_i^{desired}, z_i^{real}$: are respectively the desired output and the actual output.
The typical number of neurons in the hidden layer helps to speed up the learning

convergence. As the number of neurons in the hidden layer increased, the convergence speed decreased. A lower number of neurons gives poor generalization, but a faster convergence speed and lower computational cost.
However, in most Feed-forward Artificial Neural Networks, the number of neurons in the hidden layer is usually not known beforehand. Hence, to build a non-parametric Neural Network architecture, it is desirable to automatically identify the number of neurons. For these reasons, we used the fuzzy consensus clustering based on Iterative Pairwise Consensus or IPC. By running IPC-based fuzzy k-medoid clustering [18] on the training set, we used each medoid as a neuron prototype.
The fuzzy k-medoids is an efficient partitional clustering algorithm used for cluster analysis in Data Mining and Machine Learning. Each medoid minimizes the average distance to all patterns in the cluster. Hence, the optimal number of medoids should minimize a loss function as follows:

$$\begin{cases} O_{\pi^*} = \underset{W,Z}{\arg\min} \ = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ji}^{m} d(v_i, m_j) \\ 0 \leq u_{ji} \leq 1, \ \forall ij, \ 0 \leq i \leq n, \ 0 \leq j \leq k \\ \sum_{j=1}^{k} u_{ji} = 1, \forall i, \ \ 1 \leq i \leq n \\ 0 < \sum_{i=1}^{n} u_{ji} < n, \forall j, \ 1 \leq j \leq k \end{cases} \tag{5}$$

Where $W$ is an $n \times k$ fuzzy matrix containing the membership degree, $n$ is the number of patterns in the data set, $k$ is the number of clusters, $Q = \{m_j\}_{j=1}^{k}$ is a set of medoids.
The fuzzy k-medoids algorithm updates the membership degree $u_{ij}$ by the following formula:

$$u_{ij}^{m} = \frac{(\frac{1}{d(v_i, m_j)})^{\frac{1}{m-1}}}{\sum_{i=1}^{k} \left( \frac{1}{d(v_i, m_i)} \right)^{\frac{1}{m-1}}} \tag{6}$$

To generate a stable model by averaging the recognition accuracies generated by re-sampling-based procedure, we used the model selection technique based on Cross-Validation. The selected model is used to predict the outputs for the test samples. Each input pattern from the testing set is assigned to the category with the highest output score.

In the next section, a general description of our benchmarking model for fuzzy consensus clustering, pattern categorization and model selection is presented.

## 4 Experimental Study

The objective of our experiment is to measure the generalization ability of our Deep Learning model.

### 4.1 Data set

The data used in this study comes from a repository of Artificial and Machine Learning, which is the most widely used test collection for the diagnosis of breast cancer [8]. The features represent a range of characteristics for patients diagnosed with Breast Cancer. Each category indicates the type of cancer {M for Malignant, B for Benign}. The average length of categories $\bar{L}$ in terms of patterns is equal to 284.50, and standard deviation $\sigma_L$ is 72.50. The number of patterns across categories is highly balanced.
As shown in Figure 3, from 596 patterns representing cancerous samples, we select the first pattern from each category, and we plot their feature values. The X-axis represents the target feature, and the Y-axis shows its value.

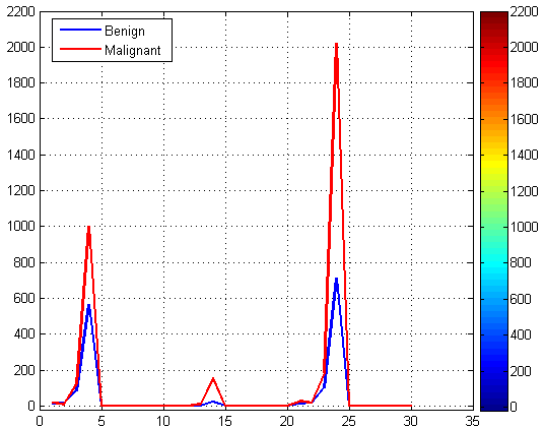


**Fig. 3.** Data distribution.

The training set and testing set are sampled from the original data set. Thirty percent of the data are selected to test the model (no theoretical justication for this percentage).

### 4.2 Pretreatment

The impact of noise on generalization is determined by the noise level in the Data Sets. It is necessary to delete all occurrences of punctuation tokens and labels for columns: "id", "diagnosis", "radius_mean", "texture_mean", etc.

### 4.3 Configuration

Our proposed architecture has been implemented on Eclipse Integrated Development Environment 2025-03 R 64-bit and some library functions such as Java SE Development Kit 24.0.1, Java Matrix Package or JAMA, etc.

### 4.4 Consensus clustering

The main goal of consensus clustering is to yield a consolidated model by combining multiple runs of the clustering algorithm.
Mathematically speaking, let $D = \{p_i\}_{i=1}^{n}$ a set of pattens, $\pi_i = \{C_i\}_{j=1}^{k}$ is a clustering consisting of $k$ clusters and $\Pi_k = \{\pi_k\}_{j=1}^{k}$ is the set of $k$ clustering of $D$, $\Pi$ the set of all possible clustering.
The aggregate models $\pi_i$ are consolidated into a single clustering $\pi^* \in \Pi_k$, which gives the best fitting of training data.
The obtained consensus clustering $\pi^*$ should minimize the following objective function:

$$\pi^* = \underset{\Pi_k \in \Pi}{\operatorname{argmin}} = \sum_{i=1}^{k} \varphi(\Pi, \Pi_{kj}) \tag{7}$$

Where $\varphi$ is a distance measure between clustering sets.

There is a wide variety of distance measures that have been used for consensus clustering. Among them there is a class of Counting pairs (e.g. Pearson's chi-squared, odds ratio (OR), pearson correlation coefficient, Fowlkes-Mallows, Mirkin distance, Rand index, Jaccard coefficient, etc.), Set matching (e.g. F-measure, Purity Dongen, and Inverse Purity, etc.), Information Theory (e.g. information gain or IG, V-measure, Mutual Information (MI), Variation of Information, Utility Function, etc.), Kernel measures (e.g. Subset Significance, Graph Kernel, etc.)[21],[10], etc.
In our Deep Learning model, we used Iterative Pairwise Consensus-based fuzzy k-medoid clustering algorithm to yield stable and robust clustering. The pairwise measure is applied to improve the clustering results, which is defined as follows,

$$S_{ij} = S(p_i, p_j) = \frac{1}{k} \sum_{p=1}^{k} \Phi(\pi_p(p_i), \pi_p(p_j)) \tag{8}$$

$$\Phi(\pi_i(p_i), \pi_i(p_j)) = \begin{cases} 1, \pi_i(p_i) = \pi_i(p_j) \\ 0, otherwise \end{cases}$$

This measure defines an $n \times n$ symmetric matrix with positive entries $S_{ij} \geq 0, S_{ij} = S_{ji}\ \forall i \neq j,\ 1 \leq i, j \leq n$, in which each off-diagonal entry is the number of shared clusterings to the total number of clustering in the set. The entries on the main diagonal are zero, i.e., $S_{ii} = 0, \quad \forall i,\ 1 \leq i \leq n$.

The average value of the silhouette index is used to measure the cluster compactness and separation, defined as follows,

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n} s(p_i) \tag{9}$$

For pattern $p_i$, the silhouette-index $s(p_i)$ is defined as follow:

$$s(p_i) = \frac{b(p_i) - a(p_i)}{max(b(p_i) - a(p_i))} \tag{10}$$

$$a(p_i) = \frac{1}{\mid C_j \mid -1}\sum_{p_i^{'} \in C_j} d(p_i, p_i^{'}), \forall p_i, p_i^{'} \in C_j, p_i \neq p_i^{'} \tag{11}$$

$$b(p_i) = \min_{C_j \neq C_k} d(p_i, C_k) \tag{12}$$

$$d(p_i, C_k) = \frac{1}{\mid C_k \mid}\sum_{p_i^{'} \in C_k} d(p_i, p_i^{'}) \tag{13}$$

Where $a(p_i)$ is the average of the distances between $p_i \in C_k$ and all patterns belonging to the same cluster, $b(p_i)$ is the average of the distances between $p_i$ and all patterns within the closest cluster.

The parameter tuning of fuzzy k-medoids has a great influence on the convergence speed of learning. Typically, the architecture configuration is described in Table. 1. The fuzzifier parameter (or fuzzy partition matrix) $m \in ]1, \infty[$ varies the level of overlapping. If $m \to \infty$, the degrees of membership become completely fuzzy. Inversely, if $m \to 1$, the memberships $u_{ij} \to 0$ or $u_{ij} \to 1$, resulting a crisp clustering. The typical value of the fuzzifier parameter is equal to 2. The threshold $\delta$ controls the speed of convergence. If we kept a high threshold value, we would meet the neighborhood vicinity of the optimal or near-optimal solution quickly. A lower threshold value leads to a more precise solution, but slower convergence and higher computational cost. The typical value of threshold parameter is equal to 0.001. The membership degree is initialized to a relatively low random value $u_{ij} \in ]0, 1]$.

**Table 1.** FCM Architecture configuration

| Parameter | Allowable value | Typical value |
|---|---|---|
| $\delta$ | ]0,1[ | 0.001 |
| $m$ | ]1,∞[ | 2 |
| $u_{ij}$ | ]0,1] | Random ]0,1] |

Clustering performance is based on maximizing the average value of the silhouette-index.
As shown in Figure 4, for each of the nine identified clusters, we select the first 30 patterns and we plot their silhouette-index values. We see that the values of IPC-based fuzzy k-medoid silhouette-index are significantly higher than the k-medoid values. The average value of the silhouette-index of IPC-based fuzzy k-medoid is equal to 0.69, and fuzzy k-medoid clustering indicates a lower value equal to 0.53.
The X-axis enumerates the chosen patterns $\{p_i\}_{i=1}^{30} \in \{C_i\}_{i=1}^{9}$. The Y-axis corresponds to the silhouette-index values.

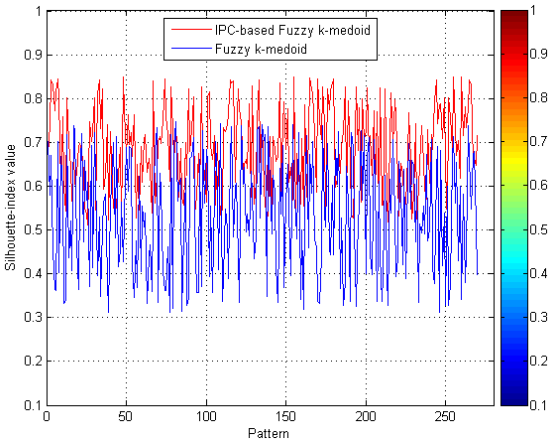


**Fig. 4.** Fuzzy k-medoid v.s IPC-based Fuzzy k-medoid clustering performance.

According to the obtained results, the Iterative Pairwise Consensus-based fuzzy k-medoid clustering algorithm yields better performance compared to fuzzy k-medoid, i.e., better clustering fuzziness, and the resulting clusters are appropriately compact and well-separated.

### 4.5 Model selection

To evaluate our Deep Learning model for pattern categorization, we used the model selection technique based on 10-fold Cross-Validation.
The data set is divided into $k$ blocks; the $k-1$ blocks are used to train the model, and the remaining block is used as test data.
We used recognition accuracy measures to evaluate the learning performance of the categorization models.
The recognition accuracy of the $i^{th}$ categorization model is computed as the number of recognized patterns divided by the total number of tested patterns,

as described by the following formula.

$$accuracy_i = \frac{\#\ recognized\ pattern}{\#\ presented\ pattern} \times 100 \tag{14}$$

However, by a re-sampling-based procedure, we used the Average Accuracy (AC) as a consistent measure for aggregating predictions, which has been adopted for Bootstrap Aggregation scheme in Deep Learning and Data Mining.
The average accuracy (AC) is aggregated as an average of the recognition accuracies over 10 estimates.

$$AC = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{c}\sum_{j=1}^{c}(\frac{\#\ recognized\ pattern}{\#presented\ pattern}) \times 100 \tag{15}$$

Where,
$k$: is the number of folds.
$c$: number of categorization models.
Figure 5 depicts the convergence of Deep Learning algorithms. The Mean Squared Deviation (MSD) is used as a tracking error to measure the difference between the desired output $y_i^d$ and actual output $y_i^a$. The goal of our learning scheme is to adjust $W$ and $W'$ matrices to find a better fit of the Data. The X-axis indicates the number of epochs or presentations of the full training set, and the Y-axis shows the amount of error. As the epochs proceed, the error oscillates, skipping over the local minimum, and the actual output gets closer to the desired output. The recognition accuracy of our Deep Learning model is equal to 97.01% after 77 iterations. The Radial Basis Function Neural Network learns after 100 iterations with a recognition accuracy equal to 96.73%.

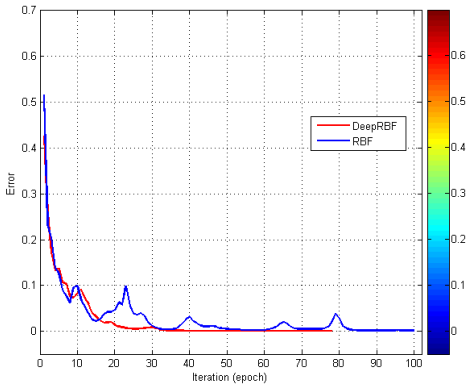


**Fig. 5.** Iteration v.s Error.

Experiments show that our typical initialization-based Deep Learning model has good performance and lower computational cost, which provides an efficient mining model for pattern categorization.

### 4.6 Evaluation

Four Deep Learning models were evaluated on the task of pattern categorization. They correspond to MultiLayer Perceptron Neural Network (MLP), Radial Basis Function (RBF), Fuzzy Adaptive Resonance Theory of Analog Multidimensional Maps (fuzzy ARTMAP), and our Deep Learning model (DeepRBF).

- *The MultiLayer Perceptron (MLP)*: is a feed-forward Neural Network architecture for supervised learning of analog data. The architecture consists of three layers of neurons: the input layer, the hidden layer, and the output layer. The back-propagation algorithm based on Stochastic Gradient Descent (or SGD) is used to update the synaptic weights, as defined by the following formulas:
  Output layer: $w_{kj} \leftarrow w_{kj} - \eta\delta_k y_j$
  Hidden layer: $w_{kj} \leftarrow w_{kj} - \eta\delta_i x_i$
  where,
  $\delta$:error.
  $x_i$: input neurone.
  $y_i$: actual output.
  $\eta$: the learning rate.
  The MLP architecture configuration needs careful tuning of parameters, particularly the weight updating scheme and learning rate.
  We applied the Stochastic Gradient Descent (SGD) algorithm as a weight updating mechanism, and we set the learning rate $\eta = 0.001$.

- *Fuzzy Adaptive Resonance Theory of Analog Multidimensional Maps*: is a competitive supervised Neural Network classifier, which is used to recognize analog and binary patterns. The competitive learning mechanism is used to avoid the stability-plasticity dilemma. The neural architecture contains two layers, namely, the input layer and the output layer, which interact with the bottom-up $b_{ij}$ and top-down $t_{ij}$ processes of long-term memory or LTM. The winning neuron is selected by applying the intersection of two fuzzy sets, as defined by the following formula:

$$J = \underset{j}{\text{argmax}} \frac{\sum_{i=1}^{2n} min(b_{ij}, x_{ji})}{\alpha + \sum_{i=1}^{2n} b_{ij}}$$

  The selected neuron is subjected to a resonance test by using the vigilance parameter.
  A more detailed study of the neural architecture configuration and parameters initialization can be found in the paper [5].

To validate the learning performance, we used *Precision*, *Recall*, and $F-measure$ indexes, which are calculated for all $k$ categories.

$$precision_{\mu} = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fp_i} \tag{16}$$

$$recall_{\mu} = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fn_i} \tag{17}$$

$F-measure_{\mu}$ measures the tradeoff between precision and recall, i.e.,

$$F-measure_{\mu} = 2 \times \frac{precision_{\mu} \times recall_{\mu}}{precision_{\mu} + recall_{\mu}} \tag{18}$$

where $TP$, $FP$ and $FN$ are True Positive, False Positive, and False Negative, respectively.

**Table 2.** The effectiveness of categorization.

| Model | Precision | Recall | F-measure |
|---|---|---|---|
| $MLP$ | 82.09 | 82.11 | 82.10 |
| $F.ARTMAP$ | 85.11 | 85.17 | 85.14 |
| $RBF$ | 84.01 | 84.15 | 84.08 |
| $DeepRBF$ | 87.17 | 85.77 | 86.46 |

According to the statistical evaluation measures, our categorization model shows greater generalization ability and enhanced recognition accuracy. In addition, our typical initialization scheme reduces the computation time and improves the convergence speed. It achieves the optimal solution and avoids local minima.
The overall conclusion is that the application of consensus clustering and model selection has a positive impact on learning stability and recognition accuracy.

## 5 Conclusion

In this paper, we presented a new scalable scheme for Radial Basis Function Network tuning using fuzzy consensus clustering and model selection. To find the number of neurons in the RBF layer, we used an IPC-based fuzzy k-medoid clustering. The IPC-based fuzzy k-medoid consensus indicates a better clustering fuzziness than fuzzy k-medoid. The typical initialization scheme is used for speeding up the learning convergence, achieving the neighborhood vicinity of the optimal solution, and thus avoiding local minima.
We used the model selection technique based on k-fold Cross-Validation to find a single consolidated categorization model for better data fitting.

Our Deep Learning model leads to an increased performance in pattern categorization and enables the discovery of complex decision boundaries. Good comparisons with the experimental studies demonstrate the multidisciplinary applications of our Deep Learning model.

## 6 Future works

Our next work is on Deep Learning, where the goal is to produce a nonparametric RBF architecture by using boosting, ensemble learning, and model selection techniques.

## Acknowledgment

We would like to gratefully acknowledge the support of the University of Quebec at Rimouski for funding this research.

## References

1. Babitha, L., Manjunatha, B., Mohammad, Q., Singh, N., Shrivastava, A., Khan, I.: Cross-domain ontology synthesis for semantic web integration: A neural network approach. In: 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE). pp. 737–742. IEEE (2024)
2. Carriero, A., Groenhoff, L., Vologina, E., Basile, P., Albera, M.: Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024. Diagnostics **14**(8), 848 (2024)
3. Chen, J., Pan, Teng andZhu, Z., Liu, L., Zhao, N.: A deep learning-based multimodal medical imaging model for breast cancer screening. Scientific Reports **14696**, 15 (2025)
4. Dang, T., Nguyen, T.T., McCall, J., Elyan, E., Moreno-García, C.F.: Two-layer ensemble of deep learning models for medical image segmentation. Cognitive Computation **16**(3), 1141–1160 (2024)
5. Djellali, C., Adda, M.: A new categorization numerical scheme for mobile robotic computing using odor data-set recognition as a case. Procedia Computer Science **94**, 199–206 (2016)
6. Dsouza, K.J., Ansari, Z.A.: Histopathology image classification using hybrid parallel structured deep-cnn models. Applied Computer Science **18**(1), 20–36 (2022)
7. Ekman, T., Barakat, A., Heiberg, E.: Generalizable deep learning framework for 3d medical image segmentation using limited training data. 3D Printing in Medicine **11**(1), 1–18 (2025)
8. Gupta, P., Garg, S.: Breast cancer prediction using varying parameters of machine learning models. Procedia Computer Science **171**, 593–601 (2020)
9. Karakostas, B.: Control of autonomous uav using an onboard lstm neural network. Journal of Ubiquitous Systems & Pervasive Networks **18**(1), 09–14 (2023)
10. Nguyen, N., Caruana, R.: Consensus clusterings. In: Seventh IEEE international conference on data mining (ICDM 2007). pp. 607–612. IEEE (2007)

11. Pathak, D., Kashyap, R.: Electroencephalogram-based deep learning framework for the proposed solution of e-learning challenges and limitations. International Journal of Intelligent Information and Database Systems **15**(3), 295–310 (2022)
12. Qu, F., Tian, E., Zhao, X.: Chance-constrained $h\_\{\}$ state estimation for recursive neural networks under deception attacks and energy constraints: The finite-horizon case. IEEE Transactions on Neural Networks and Learning Systems (2022)
13. Sharma, A., Arya, A.: Document information retrieval with deep learning. In: 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE). pp. 1–4. IEEE (2024)
14. Shoukat, M.U., Yan, L., Deng, D., Imtiaz, M., Safdar, M., Nawaz, S.A.: Cognitive robotics: Deep learning approaches for trajectory and motion control in complex environment. Advanced Engineering Informatics **60**, 102370 (2024)
15. Smyl, S., Dudek, G., Pełka, P.: Es-drnn: a hybrid exponential smoothing and dilated recurrent neural network model for short-term load forecasting. IEEE Transactions on Neural N and Learning Systems (2023)
16. Sohrabi, P., Jodeiri Shokri, B., Dehghani, H.: Predicting coal price using time series methods and combination of radial basis function (rbf) neural network with time series. Mineral Economics **36**(2), 207–216 (2023)
17. Song, J.: Research on financing risk assessment and optimization of digital economy enterprises combined with deep learning technology. In: SHS Web of Conferences. vol. 213, p. 01019. EDP Sciences (2025)
18. Tasia, E.T.E., et al.: Perbandingan algoritma k-means dan k-medoids untuk clustering daerah rawan banjir di kabupaten rokan hilir: Comparison of k-means and k-medoid algorithms for clustering of flood-prone areas in rokan hilir district. Indonesian Journal of Informatic Research and Software Engineering (IJIRSE) **3**(1), 65–73 (2023)
19. Thakur, G.K., Thakur, A., Kulkarni, S., Khan, N., Khan, S.: Deep learning approaches for medical image analysis and diagnosis. Cureus **16**(5) (2024)
20. Tian, J., Li, H., Qi, Y., Wang, X., Feng, Y.: Intelligent medical detection and diagnosis assisted by deep learning. Applied and Computational Engineering **64**, 120–125 (2024)
21. Vega-Pons, S., Ruiz-Shulcloper, J.: A survey of clustering ensemble algorithms. International Journal of Pattern Recognition and Artificial Intelligence **25**(03), 337–372 (2011)
22. Victor, J.O., Chew, X., Khaw, K.W., Chong, Z.L.: Cratsm: An effective hybridization of deep neural models for customer retention prediction in the telecom industry: Manuscript received: 8 april 2024, accepted: 5 june 2024, published: 15 september 2024, orcid: 0000-0001-5539-1959, https://doi. org/10.33093/jetap. 2024.6. 2.10. Journal of Engineering Technology and Applied Physics **6**(2), 66–78 (2024)
23. Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., Manners, D.N., Tonon, C.: Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives. Diagnostics **12**(8), 1850 (2022)
24. Xue, Y., Tong, Y., Neri, F.: An ensemble of differential evolution and adam for training feed-forward neural networks. Information Sciences **608**, 453–471 (2022)