International Journal of Information Retrieval Research
Volume 12 • Issue 1

A New Scalable Deep Learning Model of Pattern Recognition for Medical Diagnosis Using Model Aggregation and Model Selection

Choukri Djellali, University of Quebec at Rimouski, Canada*
Mehdi Adda, University of Quebec at Rimouski, Canada

ABSTRACT

In recent years, pattern recognition has become a research area with increasing importance using several techniques. One of the most common techniques used is deep learning. This paper presents a new deep learning model to pattern recognition for medical diagnosis. The uncovering of hidden structures is performed by feature selection, model aggregation, and model selection. The deep learning model has the ability to reach the optimal solution and create complex decision boundaries when used to look for and diagnose breast cancer. The evaluation, based on 10-fold cross-validation, showed that the proposed model, which is named BaggingSMF, yielded good results and performed better than radial basis function, bidirectional associative memory, and ELMAN neural networks. Experimental studies demonstrate the multidisciplinary applications of the model.

KEYWORDS

Artificial Neural Networks, Data Mining, Deep Learning, Feature Selection, Fuzzy Logic, Medical Diagnosis, Model Aggregation, Model Selection, Pattern Recognition

1. INTRODUCTION

Pattern recognition has been widely used in Data Mining community, significantly improving the state-of-the-art models. It is the process of extracting hidden pattern and regularities from a large data set to predict outputs.

It is defined as the study of how machines can observe the environment, learn to distinguish various patterns of interest from its background, and make reasonable decisions about the categories of the patterns. During recognition, the given objects are assigned to a prescribed category (Kaur & Sharma, 2013).

Pattern recognition has been successfully applied to a wide range of application domains, including Face Recognition (Hu et al., 2015; Wen et al., 2016), Speech Recognition (Deng et al., 2013), Medical Diagnosis (Yamins & DiCarlo, 2016), Computer Vision (Kendall & Gal, 2017), Biology (Angermueller

DOI: 10.4018/IJIRR.316131 *Corresponding Author

et al., 2016), Video Recognition (Kahou et al., 2016; Ng et al., 2015), Image Compression (Cheng et al., 2018), Network Security (Yuan et al., 2014), Gesture Detection and Localization (Neverova et al., 2014), and many other domains.

Recently, Deep Learning has proven to be successful and efficient in identifying the complex decision boundaries for pattern recognition. It is used widely in machine learning, for modeling high-level abstractions in data using different types of nonlinear transformations.

In deep learning, there are two main types of training: supervised, and unsupervised. In supervised learning mechanism, the deep learning model learns from labeled training dataset. Unsupervised learning looks for correlated patterns in a data set to infer the hidden structure.

Most pattern recognition algorithms based on Deep Learning are sensitive to noise, Bellman's curse of dimensionality, architecture configuration and instability recognition.

On one hand, model aggregation is a powerful technique that combines multiple pattern recognition algorithms to improve the stability and accuracy.

On the other hand, Cross-Validation is a data-based machine learning technique applied to address bias-variance trade-off, and thus, to improve the recognition accuracy.

Based on these premises, a novel Deep Learning-based method for pattern recognition which can provide improvement in recognition accuracy, is described.

We demonstrate its ability to reach the optimum solution and obtain more cluttered decision boundaries on medical diagnosis problem.

This paper is organized as follows. In Section 2, we discuss the more commonly used Deep Learning architectures for pattern recognition, our research questions and the drawbacks of current models. The Deep Learning architecture is described in detail in Section 3. Section 4 describes the benchmarking models used for performance evaluation. Finally, Section 5 to 6 present our conclusions and future works.

The next section of the paper presents a survey of the state of the art and discusses the advantages and drawbacks of pattern recognition algorithms.

2. LITERATURE REVIEW

In recent years, the use of Deep Learning has gained popularity in pattern recognition. Several algorithms have been proposed to solve the problem of pattern recognition. These models include:

- **Artificial Neural Networks (ANNs) (also known as connectionist computational models):** Designed to simulate the biological Neural Networks. The neural architecture is composed of many interconnected units usually known as artificial neurons. It is widely used for more complex tasks such as categorization, prediction, clustering, regression, and summarization, etc. Among the most popular models in this category, we quote Self Organizing Map (SOM), Growing Neural Gas (GNG), Adaptive Resonance Theory (ART), Real-Time Recurrent Learning (RTRL), Gated Recurrent Units (GRUs), Boltzmann Machine, Learning Vector Quantization (LVQ), Deep Belief Networks (DBNs), Hopfield, Bidirectional Associative Memory (BAM), Growing Cell Structures (GCS), Recurrent Convolutional Neural Network (RCNN), etc. (Duda et al., 2012; Liang & Hu, 2015).
- **Particle Swarm Optimization (PSO) models:** It is an optimization technique introduced by Dr. Eberhart and Dr. Kennedy in 1995, inspired by collective intelligence. It is basically a heuristic search method using an iteration strategy based on the best-known solution found by all particles in the search space. The most common models in this category are Fuzzy Adaptive Particle Swarm Optimization, PSO-GA, GAI-PSO, Differential Evolution-PSO (DEPSO), DEPSO-Scout, ICSPSO, etc. (Sengupta et al., 2018).
- **Genetic algorithms (GAs):** Are computing models inspired by the theory of natural evolution. The selection process based on the fitness of each chromosome is used to create off-springs

for the next generation. Crossover and mutation operators are applied to provide high genetic diversity. We find in this category the following models: Multiple Objective Genetic Algorithms (MOGAs), Multiple-deme GA, Coarse-grained parallel GA, ASPARAGOS, multiple-population GA, master-slave GA, etc. (Cantu-Paz, 1998).

- **Simulated annealing:** Inspired by the annealing process in physical metallurgy, in which the heat treatment process alters the properties and metallurgical structures of metals. We quote in this category the following models: Chaos Simulated Annealing (CSA), MOSA, SMOSA, UMOSA, Pareto Simulated Annealing (PSA), WMOSA, etc. (Suman & Kumar, 2006).

Pattern recognition has been significantly advanced through the use of Deep Learning techniques. Many of the recent advancements have led to the appearance of several approaches for pattern recognition.

Albeahdili et al. (2015) proposed a new Deep Learning model based on Convolutional Neural Network (CNN) for image recognition. The CNN architecture contains three layers, namely 96C-96S-256C-256S-96C-96S-90F-120F-x-softmax, where C is a Convolution layer, S is a sub-sampling layer, and F is a full conned layer. The output or Soft max layer is used for performance evaluation. This approach is applied on MNIST, CIFAR-10, and CIFAR-100 data sets. The evaluation showed that the proposed model yielded good results and performed better in comparison with the state of the art.

Tarik et al. (2014) presented an Artificial Neural Network (ANN) based on MultiLayer Perceptron (MLP) for digital movement images restoring. A supervised learning method based on a Back-propagation algorithm is used for modifying the connection weights between neurons. To find the typical topology, this approach used three different architectures based on parameter tuning. The Neural Network architecture is composed of 64 neurons in the input layer, one hidden layer of 4 neurons and an output layer of 64 neurons. The experiment results showed an improvement of recognition accuracy of 96%.

Cheng et al. (2018) introduced new Deep Learning based on Convolutional AutoEncoder (CAE) for image compression. For training, this model uses a rate-distortion loss function. Principal Components Analysis (PCA) is used to rotate the representation space and increase the speed of training. The encoder architecture contains the pre-processing steps, CAE computation, PCA rotation, quantization, and entropy coder. The decoder part is used to mirror the encoder architecture. Experimental studies based on the Kodak data set, showed that the Deep Convolutional AutoEncoder performed well than JPEG and JPEG2000 in terms of PSNR, and achieved 13.7% BD-rate decrement compared to JPEG2000.

Liang et al. (2015) proposed a Deep Learning model based on Recurrent Convolutional Neural Network (RCNN) for object recognition. RCNN neural architecture contains a stack of RCLs interleaved with max-pooling layers. The Deep Learning model is trained using the Back Propagation Through Time (BPTT) and Stochastic Gradient Descent (SGD) algorithms. This approach is applied to four benchmark object recognition datasets: CIFAR-10, CIFAR-100, MNIST, and SVHN. The evaluation showed that the proposed model yielded good results and performed better than the state-of-the-art models.

Yaniv et al. (2015) presented a new system for medical application of chest pathology detection in x-rays, by using Convolutional Neural Networks (ConvNets or CNNs). The neural architecture is a combination of DeCAF5 and PiCoDes. The benchmarking Dataset contains 93 frontal chest x-ray images that comes from Sheba Medical Center. The evaluation based on leave-one-out-Cross-Validation, showed that the proposed model achieved the best recognition accuracy compared to the state-of-the-art models.

Recent works, Wentao et al. (2016) proposed an end-to-end fully connected deep LSTM network for skeleton-based action recognition. The LSTM Neural Network consists of three LSTM layers and two feed-forward layers. The Dropout technique is used to enhance the model generalization. This deep Neural Network is applied on SBU Kinect Interaction, HDM05 and CMU datasets. Experimental

results based on 10-fold Cross-Validation showed that the proposed architecture outperformed existing Deep Learning models.

Most pattern recognition models based on Deep Learning are perturbed by noisy features, architecture configuration, and learning instability. Moreover, the decision boundaries yielded are not well separated.

In this present study, we present a new Deep Learning model for pattern recognition using features selection, an aggregation scheme, and model selection.

In the next section, we introduce the architecture adopted in our approach.

3. ARCHITECTURE OF OUR DEEP LEARNING MODEL

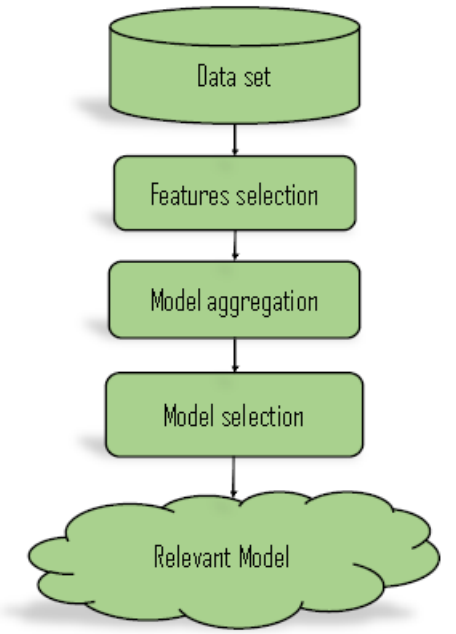
As shown in Figure 1, our Deep Learning model starts by feeding recognition algorithm from a large data set. The presence of irrelevant features adds noises into decision regions separated by decision boundaries, maximizes the number of misclassification, and thus, negatively impacts the performance. Therefore, features selection step is applied before input patterns passed into the Deep Learning model.

To categorize patterns according to their contents, we used a Bootstrap aggregation scheme based on MultiLayer Perceptron (MLP), Fuzzy Self Organizing Map (SOM) and Fuzzy Adaptive Resonance Theory of Analog Multidimensional Maps (Fuzzy ARTMAP). This meta-modeling technique improves the stability and accuracy of pattern recognition algorithms.

To minimize the variance and bias of our model, we used an effective sampling technique based on k-fold Cross-Validation. The data set is sampled into a training set and a testing set. Of the k blocks, a single block is retained as the test data for performance evaluation, and the remaining (k-1) blocks are used as training samples. The aggregate model created from a combination of aggregated models improves the stability and accuracy.

Finally, the test sets are used for evaluating the performance of the techniques used for pattern recognition.

Figure 1. Deep Learning architecture



Through experimental studies, we first present the used data set for training, discuss the features selection in the representation space and then describe the parameters tuning for architecture configuration. We present the benchmarking models and describe the measures used for performance evaluation, and demonstrate the ability of our model to reach the optimal solution.

4. EXPERIMENTAL STUDY

4.1. Data Set

The breast cancer Wisconsin Data set that we feed into our Deep Learning model comes from the Machine Learning Repository, available online at[1]. It is the most commonly used data set for Medical diagnosis.

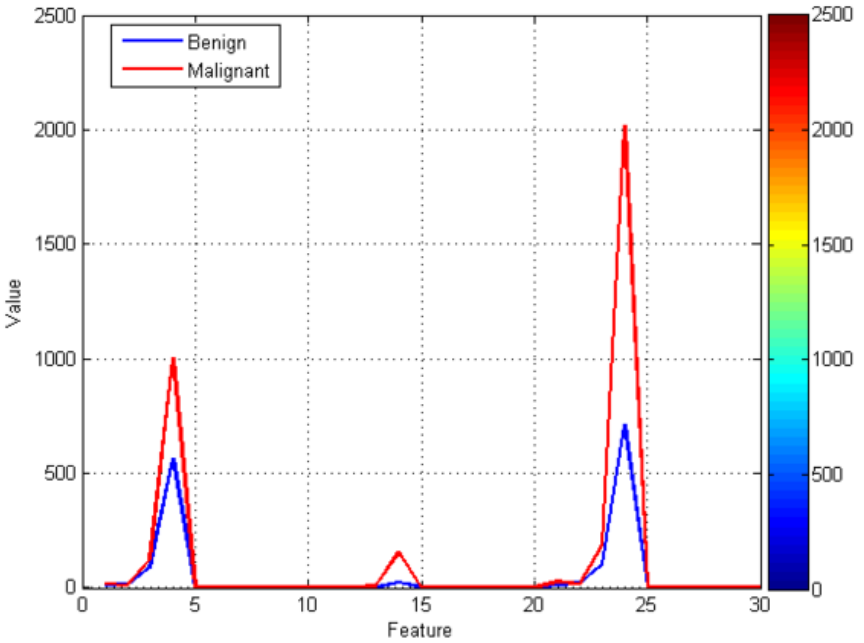
As shown in Figure 2, from 569 patterns, we select the first pattern from each category and we plot their features values. The X-axis represents 30 real-valued input features and Y-axis shows their values. Features represent characteristics of the cell nuclei present in digitized image of a fine needle aspirate (FNA) of a breast mass.

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

4.2. Configuration

Our proposed architecture has been implemented on Neon.1a Release (4.6.1) eclipse integrated development environment 64-bit and some library functions such as JDK 11.0.6 + Java EE, Java Matrix Package or JAMA[2], etc.

Figure 2. Training input patterns, Feature vs. Value



4.3. Features Selection

Figure 3 shows a histogram of the frequency distribution of features in the representation space. The values of the series are displayed as a percentage of each class {relevant features, noisy features}. Noisy features around data have 3.12% on average, while relevant features have 96.88% on average.

The features related to noise are irrelevant and do not affect the representation space, i.e., their inclusion would reduce the recognition accuracy. Therefore, the pattern recognition algorithm uses data containing only the relevant features for learning and testing.

4.4. Model Aggregation

Deep learning is a powerful technique used to learn the decision boundaries, by using multiple layers of non-linear features.

Formally, Deep Learning is an approximation of a target function ψ by a classifier ψ'' which is defined as follows:

$$\begin{cases} \psi : P \times C \mapsto \{T, F\} \approx \psi'' : P \times C \mapsto \{T, F\} \\ \quad if(\psi(p_i, c_i) = T \Rightarrow p_i \in c_i \ else \ p_i \notin c_i \end{cases}$$

The goal of Deep Learning is to select a function ψ that closely approximates a target function ψ'' by minimizing the generalization error defined by the following formula:

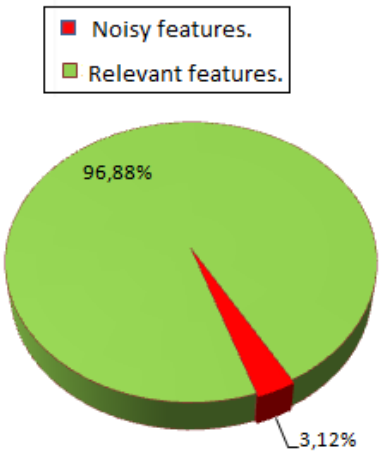
$$e = \underset{\psi''}{argmin}\left(\frac{1}{n}\sum_{i=1}^{i=n} f_L\left(\psi\left(p_i, c_i\right)\right)\right)$$

$$\forall\left(p_i, c_i\right) \in S_n$$

where $P \subset \Re^n, C \subset \Re^d$ and $S_n = \left\{\left(p_1, c_1\right), \left(p_2, c_2\right), ..., \left(p_n, c_d\right)\right\}, d \leq n$.

f_L : loss function.

Figure 3. Noisy features vs. relevant features



As previously mentioned, three Neural Networks are used in our approach as aggregated models.

- **Fuzzy Adaptive Resonance Theory of Analog Multidimensional Maps (Fuzzy ARTMAP):** It is a Neural Network that has been successfully applied to design a stable model for the plasticity-elasticity dilemma, i.e., incremental learning. As shown in Figure 4, the three-layered architecture contains two fuzzy ART Neural networks and an inter-ART module. The input layer of the Neural Network is fully interconnected to its output layer, in which the input patterns are described in complement code representation. Supervised incremental learning uses fuzzy rules to define the distance between input patterns and synaptic weighs. The match tracking mechanism is used for vigilance parameter tuning.

The neural architecture configuration needs careful tuning of parameters. Table 1 summarizes the architecture configuration with an initialization scheme. Five parameters namely ascending weights b_{ij}, backward weights t_{ij}, the vigilance ρ, the learning rate β, and the choice parameter α, are used to control the learning process.

At the start of training, the ascending weights $b_{ij}(0)$ are randomly initialized by low values, and backward weights $t_{ij}(0)$ are initialized by the value 1. The resonance (also known as vigilance) parameter ρ controls the number of hyperboxes. When the resonance value increases, the number of hyperboxes also increases, which leads to over-training. The typical value of ρ is 0.9. The parameter β controls the learning speed. A lower learning rate value leads to more precise solution, but slower training. The typical value of β is 0.9. The choice parameter α takes its values in the range $]0,\infty[$. The typical value of α is 0.9 (Decanini et al., 2011; Djellali, 2014).

Figure 4. Fuzzy ARTMAP architecture

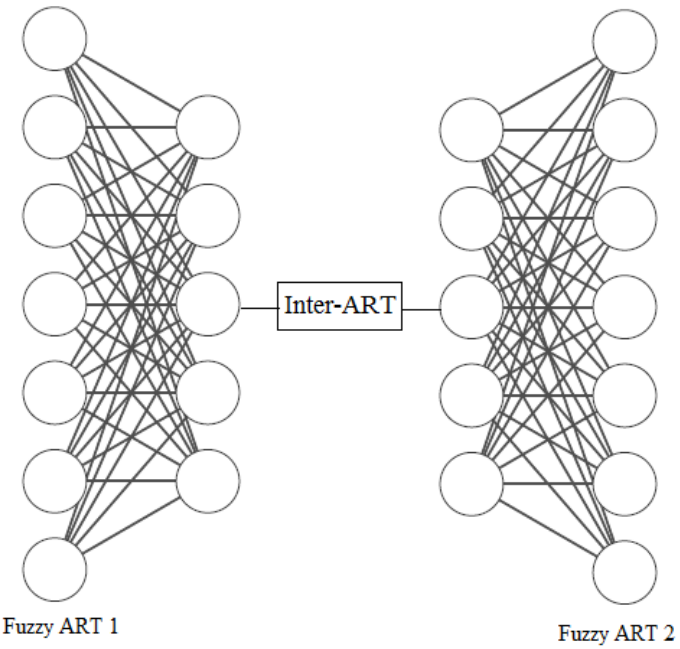


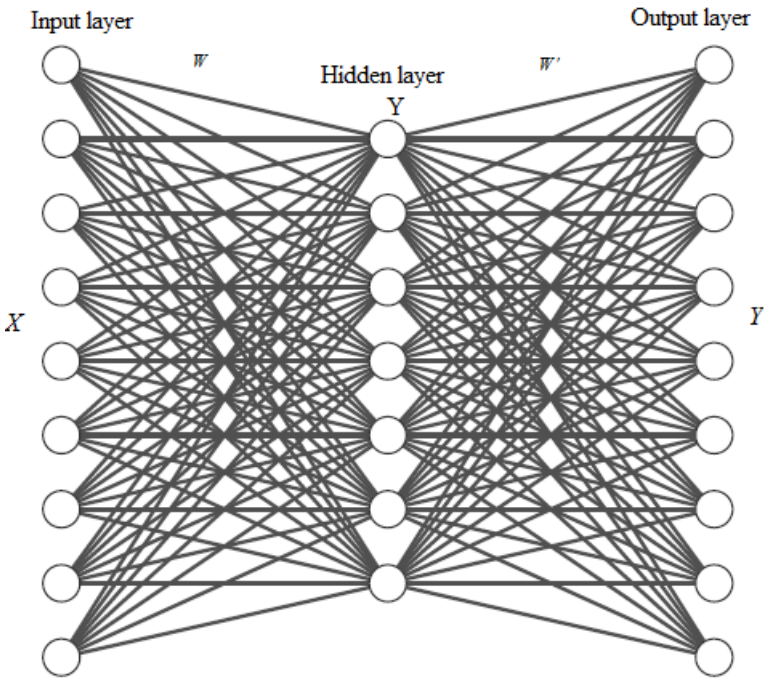
Table 1. Fuzzy ARTMAP architecture configuration

Parameter	Allowable value	Typical value
ρ	$\rho \in]0,1]$	0.9
b_{ij}	$b_{ij} \in]0,1]$	small random value
t_{ij}	$t_{ij} \in]0,1]$	1
α	$\alpha \in]0,\infty[$	0.9
β	$\beta \in]0,1]$	0.9

- **MultiLayer Perceptron (MLP):** It is a fully connected network, in which each neuron is connected to every neuron in adjacent layers. As shown in Figure 5, the neural architecture consists of an input layer, hidden layer, and output layer. The weights are calculated using the backpropagation algorithm.

Stochastic Gradient Descent (SGD) algorithm is used for training (Duda et al., 2012; Kahou et al., 2016). The rules for modifying synaptic weights are defined as follows:

Figure 5. MultiLayer Perceptron architecture



- Output layer:

$$w_{kj}^{'} \leftarrow w_{kj}^{'} - \eta \delta_k y_j$$

- Hidden layer:

$$w_{kj} \leftarrow w_{kj} - \eta \delta_k x_i$$

where:

δ : error

x_i : input neurone

y_j : actual output

η : the learning rate

Parameters tuning has a positive impact on MLP training speed. We used the number of hidden layers, learning rate η, and the number of neurons in the hidden layer as three parameters to control the training speed. The Neural architecture 10-7-2 (3 neural layers, network with 10 inputs and 2 outputs, 1 Hidden layer with 7 neurons) is used for MLP configuration. We set the learning rate $\eta = 0.001$.

- **Self Organizing Map (also known as topology preserving maps):** Is an Artificial Neural Network that contains an input layer and output layer as shown in Figure 6.

The winner neuron (also known as Best Matching Unit, BMU) is selected by the following criterion:

$$i\left(X\right) = \underset{j}{argmin}\, X - W_j$$

Unsupervised training based on a competitive mechanism is used to modify the connection weights between neurons, in which the winner neuron and its neighboring neurons update their weights as follow:

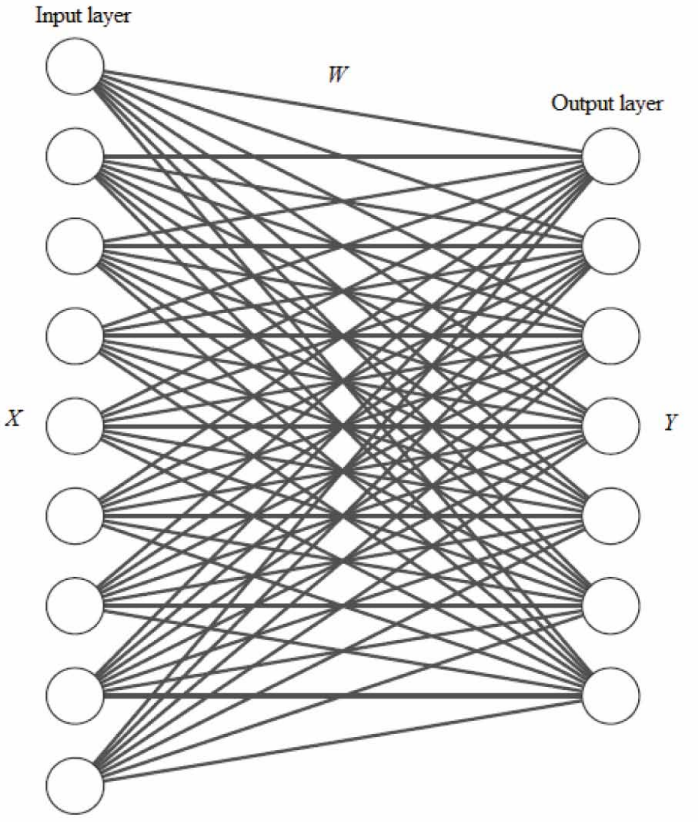
$$W_j\left(t+1\right) \leftarrow W_j\left(t\right) + h_{ji(x)}\left(t\right)\left(X - W_j\left(t\right)\right)$$

where $X(x_1, x_2, ..., x_m)^T$, $W_j\left(w_{j1}, w_{j2}, ..., w_{jm}\right)$, $\forall j, 1 \le j \le l$ and $h\left(t\right)$ represent m-dimensional input pattern, weight vector of neuron *j* in the output layer and neighborhood function respectively and l is the number of neurons in the output layer.

For parameters tuning, we used decreasing Gaussian function as a neighborhood function of the winner neuron:

$$h_{ji(x)}^{c}\left(t\right) = \alpha\left(t\right) e^{\left(\frac{-R_c - R_{ij}^{\ 2}}{2(\eta_{ij}^{c}(t))^2}\right)^2}$$

Figure 6. SOM architecture



The radius of the neighborhood is defined as follows:

$$\eta_{ij}^{c}\left(t\right)=\eta\left(0\right)e^{\left(-\frac{t}{T}\right)}$$

where $\eta\left(0\right)$ is the initial value of the radius of the neighborhood.

The learning rate is defined as an inverse of time:

$$\alpha\left(t\right)=\alpha\left(0\right)\frac{1}{t}$$

where T and t are the number of iterations and order number of a current iteration respectively.

Table 2 lists parameters tuning for SOM configuration. The initial weights W_j are initialized to small Random values. The number of clusters to be formed l is 2. The initial value of the radius of the neighbourhood $\eta\left(0\right)$ is 1 and the number of iterations is 200. The initial value of learning rate $\alpha\left(0\right)$ is 0.6.

Table 2. SOM Neural Network configuration

Parameter	Allowable value	Typical value
T	$\left[1,\infty\right[$	200
$\eta\left(t\right)$	$\left]0,1\right]$	1
W_j	$\left]0,1\right]$	small random value
$\alpha\left(0\right)$	$\left]0,1\right]$	0.6
l	$\left]0,\infty\right[$	2

The typical initialization scheme is applied to improve the convergence speed of training and achieve the neighborhood vicinity of the optimal solution.

We used Mean Square Error (MSE) as a measure of how well the models fit data, which is the average squared difference between the desired outputs and current outputs.

This criterion is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{i=n}(o_i^c - o_i^d)^2$$

where o_i^c stands for current output, o_i^d is for the desired output, and n is for the number of neurons in the output layer.

Figure 7 illustrates the learning curve of Neural Networks. The general goal is to learn a decision boundary by updating the synaptic weights. X-axis represents the number of iterations and Y-axis indicates the generalization error.

In the light of the results, the training error tends to decrease monotonically as the presentation of patterns increases, it can increase or oscillate.

At the start of the learning, the curve shows a high error which indicates that the training input patterns are not well separated. After training, the curve shows a low error and this means that the learned model tends to be close to the training patters.

It is undoubtedly to see that the Deep Learning models escape from local minima to reach a global minimum.

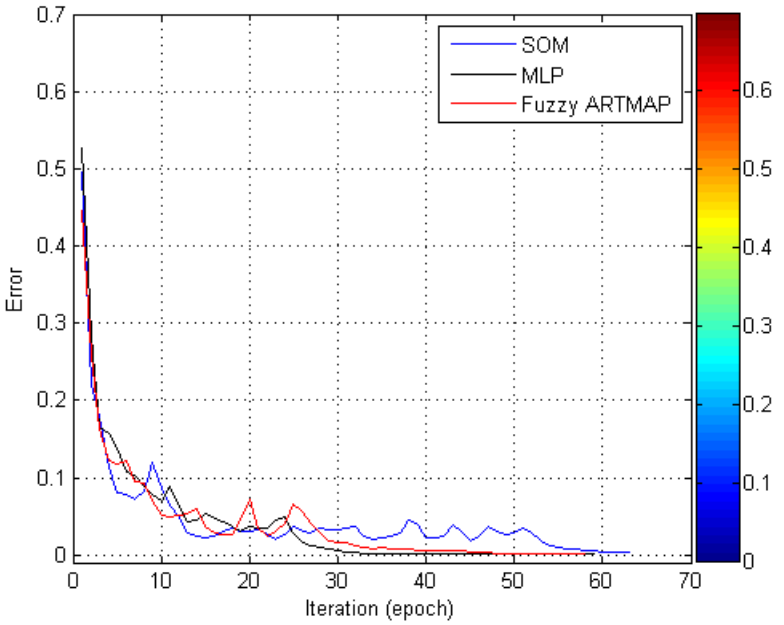
The recognition accuracy of Multilayer Perceptron is equal to 95.91% after 59 iterations. Fuzzy ARTMAP Neural Network learns after 58 iterations with recognition accuracy equal to 96.79%. Compared to MLP and Fuzzy ARTMAP, the recognition accuracy of the Self Organizing Map is equal to 95.11% with more training, i.e., the number of complete passes through the training patterns is equal to 63.

In the next section, we review the benchmarking models and performance measures used to test the effectiveness of our Deep Learning model.

4.5. Evaluation and Baseline Models

Our experimental study was designed to compare four Deep Learning models namely Radial Basis Function (RBF), Long Short Term Memory (LSTM), ELMAN and our aggregated model (BaggingSMF):

Figure 7. Learning curve, Iteration vs. Error



- **Radial Basis Function (RBF):** Is a Convolutional Neural Network (CNN) that contains an input layer, one hidden layer, and an output layer that computes the current outputs. These weights are adjusted by a supervised learning mechanism. The outputs are calculated using a non-linear RBF activation function. The feedforward mechanism is used to adjust the weights of neurons.
- **Bidirectional Associative Memory (BAM):** Is a recurrent neural architecture that learns the long-term dependencies. The neural architecture is a hetero-associative memory that maps an input layer to an output layer. The Hebbian mechanism is used for synaptic weight learning.
- **ELMAN:** Is a feedforward Neural Network architecture consisting of an input layer, a hidden layer, a context layer, and an output layer. The context layer is used to store the outputs of neurons in the hidden layer. The backpropagation algorithm is used to train the entries of the weight matrix.

The data sets were prepared as described in section (4.1).

We used a priori knowledge about the data as a direct way to validate the results.

In order to obtain stable scoring results, we applied model selection techniques based on 10-fold Cross-Validation. By running repeated 10-fold Cross-Validation on training patterns, the aggregate estimation is defined as the average of the estimations obtained on each fold.

$f-mesure$, $precision$, $recall$ measures are used for performance evaluation, which are calculated for all k folds:

$$precision_{\mu} = \frac{\sum_{i=1}^{c} tp_i}{\sum_{i=1}^{c} tp_i + fp_i}$$

$$recall_{\mu} = \frac{\sum_{i=1}^{c} tp_i}{\sum_{i=1}^{c} tp_i + fn_i}$$

$f - mesure_{\mu}$ index weights average of the $precision_{\mu}$ and $recall_{\mu}$, i.e.:

$$f - mesure_{\mu} = 2 \times \frac{precision_{\mu} \times recall_{\mu}}{precision_{\mu} + recall_{\mu}}$$

where tp, fp and fn are true positive, false positive, and false negative, respectively.

The following table lists the recognition accuracies of Deep Learning models.

According to the above statistical evaluation measures, our Deep Learning model has good performance. These results demonstrate the potential benefit of applying the aggregation techniques for pattern recognition. In addition, Cross-validation is indispensable to reduce variance and improve recognition accuracy.

The overall conclusion is that the application of aggregation and model selection techniques has a positive impact on learning stability and recognition accuracy.

5. CONCLUSION

In this paper, we presented a new Deep Learning model for pattern recognition. Our approach exploits features selection, model selection and aggregation techniques to increase the productivity of knowledge extraction. Features selection is used to reduce noise and increase training speed. Model Aggregation is applied to improve the stability and accuracy.

Parameters tuning is critical for learning performance. We used the typical initialization scheme to reduce the computation time and improve the convergence speed of training. In addition, our Deep Learning model has the ability to learn complex decision boundaries and avoid local minima. Experimental studies conducted on Medical diagnosis indicate that our proposed model outperformed existing Deep Learning models, and gives an enhanced generalization with the best recognition accuracy.

Table 3. The effectiveness of pattern recognition

Deep Learning model	$precision$	$recall$	$f - mesure$
RBF	82.07	77.65	79.80
BAM	84.15	76.98	80.40
ELMAN	85.79	79.15	82.34
BaggingSMF	87.15	82.11	84.55

FUTURE WORKS

The purpose of our next work is to improve generalization performance, i.e., choose the best model that satisfies the trade-of bias-variance. A new algorithm can be designed based on boosting and Ensemble learning techniques.

ACKNOWLEDGMENT

The authors acknowledge the support of university of Quebec at Rimouski for funding this research.

REFERENCES

Albeahdili, H., Alwzwazy, H., & Islam, N. (2015). Robust Convolutional Neural Networks for Image Recognition. *International Journal of Advanced Computer Science and Applications*, *6*(11), 105–111.

Angermueller, C., Parnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology, 12*(7), 878.

Bar, Y., Diamant, I., Wolf, L., & Greenspan, H. (2015). *Deep learning with non-medical training used for chest pathology identification. In Medical Imaging 2015: Computer Aided Diagnosis*. International Society for Optics and Photonics.

Cantu-Paz. (1998). *A survey of parallel genetic algorithms. Calculateurs paralleles, reseaux et systems repartis, 10(2), 141-171*.

Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2018). Deep convolutional autoencoder-based lossy image compression. *2018 Picture Coding Symposium (PCS),* 253-257.

Decanini, J., Tonelli-Neto, M., Malange, F., Minussi, C. (2011). Detection and classification of voltage disturbances using a fuzzy-ARTMAP-wavelet network. *Electric Power Systems Research, 81*(12), 2057-2065.

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. *IEEE International Conference on Speech and Signal Processing (ICASSP)*, 8599-8603. doi:10.1109/ICASSP.2013.6639344

Djellali, C. (2014). A new conceptual model for dynamic text clustering Using unstructured text as a case. In *Proceedings of the 2014 International conference on Computer Science Software Engineering 13*. ACM.

Duda, R., Hart, P., & Stork, D. (2012). *Pattern classification*. John Wiley & Sons.

Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., Li, S., & Hospedales, T. (2015). *When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. In Proceedings of the IEEE international conference on computer vision workshops*. IEEE.

Kahou, S., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., & Boulanger-Lewandowski, N. (2016). Emonets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User Interfaces, 10*(2), 99-111.

Kaur, P., & Sharma, M. (2013). Classification in Pattern Recognition: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*, 4.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision. *Advances in Neural Information Processing Systems*.

Liang, M., & Hu, X. (2015). *Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

Neverova, N., Wolf, C., Taylor, G., & Nebout, F. (2014). Multi-scale deep learning for gesture detection and localization. In *Workshop at the European Conference on Computer vision* (pp. 474-490). Springer.

Ng, H., Nguyen, V., Vonikakis, V., & Winkler, S. (2015). *Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction*. ACM.

Sengupta, S., Basak, S., & Peters, R. (2018). *Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives. Machine Learning and Knowledge Extraction, 1(1), 157-191*.

Suman, B., & Kumar, P. (2006). A survey of simulated annealing as a tool for single and multiobjective optimization. *Journal of the Operational Research Society, 57*(10), 1143-1160.

Tarik, H., Kodad, M., & El Miloud, J. (2014). Digital movements images restoring by artificial neural networks. *Computing in Science & Engineering*, *4*(2), 36–42.

Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision* (pp. 499-515). Springer.

Yamins, D., & DiCarlo, J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience, 19*(3), 356.

Yuan, Z., Lu, Y., Wang, Z., & Xue, Y. (2014). *Droid-sec: deep learning in android malware detection. ACM SIGCOMM Computer Communication Review, 44(4), 371-372*. doi:10.1145/2619239.2631434

Zhu, W. (2016). Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. *Thirtieth AAAI Conference on Artificial Intelligence*.

ENDNOTES

[1] https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/
[2] https://math.nist.gov/javanumerics/jama/