Classification of Brain MRIs using Improved Firefly Algorithm Based Ensemble Model

Sarada Prasanna Pati, Debahuti Mishra

Abstract: In this work we propose an automated medical image diagnosis system for diagnosing the healthy and diseased brain conditions by analyzing the brain MR images. The system uses a weighted classifier combination base ensemble model for classification of the brain images. In this approach four different base classifiers namely ANN, KNN, SVM and Naïve Bayes are first employed for training. After successful validation of each model the output obtained by each individual model is optimally weighted using unimproved Firefly algorithm as well as other established biologically inspired optimization techniques like Particle Swarm Optimization (PSO), Firefly Algorithm (FA) and Genetic Algorithm (GA) based techniques to produce the best possible results. Prior to classification task, the feature extraction for the image datasets are performed using DWT following which, PCA is used for dimensionality reduction. Finally the results obtained are compared with the results obtained by the individual classifiers as well as by the other three weighted ensemble schemes optimized using PSO, GA and AF. It is in general demonstrated that in all cases the proposed method outperforms other competitive methods.

Index Terms: Brain image classification, Classifier Ensemble, Discreet Wavelet Transformation, Principal Component Analysis, Firefly Algorithm, Improved Firefly Algorithm,

I. INTRODUCTION

Magnetic resonance (MR) imaging of brain is the most feasible and widely adopted method by radiologists and clinical practitioners for detection of normal and pathological brain conditions. In spite of the knowledge and experience, the manual interpretation and analysis of MR images by clinical experts is restricted by the human vision system. This is due to the large volume of information contained in an MR image which are hard to interpret by human vision [1] [2]. That is why use of automated image analysis methods utilising machine learning and image processing techniques have become pervasive in recent years in the field of neuroimaging.Literature shows that several automated diagnostic systems have been developed for accurate interpretation and better objective assessment of the brain MRIs and there by aiding to the accurate diagnosis of the brain diseases [3] [4] [5] [6] [7]. The steps involved in development of such automated systems primarily include feature extraction, feature reduction and classification [4]. Feature extraction in image analysis becomes imperative and refers to extraction of the relevant feature from the original image to prepare the feature vector. Discreet Wavelet Transformation (DWT) [4] [5] [6] and Gray Level Co-Occurrence Matrix (GLCM) [8] [9] are the two most popular techniques applied widely for feature extraction in brain MR image analysis [3]. In neuroimaging studies it is not often possible to collect sufficiently large number of samples and above that, the higher dimension of brain image data leads to problem, called curse of dimensionality [10]. Therefore feature reduction (also called dimensionality reduction) becomes inevitable. High dimension of feature vector may lead to design complexity of the model and may also result in over fitting/under fitting of data. Principal Component Analysis (PCA) [4] [11] [12], Independent Component Analysis (ICA) [13], Genetic Algorithm (GA) [14] are some of the preferred feature reduction techniques in case of brain image analysis. The final step is to design and train the classifier to classify the MR images of normal or pathological brains. Literature on brain image classification suggest that a wide range of machine learning techniques that are been effectively used for the purpose includes Artificial Neural Network (ANN) [4] [12], Support Vector Machines (SVM) [5] [15], k-Nearest Neighbors (k-NN) [4] [11], Classification and Regression Tree (CART) [11], Random forest [11], Convolutional Neural Network (CNN) [16], Probabilistic Neural Network (PNN) [6] [7], Fuzzy c-Means Clustering (FCM)[17] etc. to name a few. Application of ensemble of classifiers has been a relatively recent trend in machine learning to improve on the classification performance. Classifier ensemble models essentially integrate the individual performance of a variety of constituent classifiers (may be homogeneous or heterogeneous) in some way with an objective to increase the overall performance of the system. There exist several state-of-the-art methods like bagging or boosting to appropriately combine the decisions of such multiple classifier ensemble systems. There can be several approaches to combine the results of multiple classifiers based on the type of classifier output. Weighted combination model of multiple classifier systems is yet another approach well studied in several researches [18] [19] [20] [21]. The basic idea here is to find an optimal combination of weights that are assigned to the output of base classifiers based on their classification accuracy such that the whole performance of the combined system is optimized. A weighted classifier combination method is proposed in [18] where PSO is used for optimizing the weights. GA was used for weight optimization in [19], where the authors presented a classifier ensemble model for handwriting text recognition. Similarly in [20] a multiple classifier ensemble was presented for predicting train accidents where GA was used for optimally deciding the weights associated to the outcome of the base classifiers.

Revised Manuscript Received on March 10, 2019

Sarada Prasanna Pati, Department of Computer Science & Engineering, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India.
Debahuti Mishra, Department of Computer Science & Engineering, Siksha O Anusandhan Deemed to be University, Bhubaneswar, India.

Published By: Blue Eyes Intelligence Engineering & Sciences Publication
An ensemble based net asset value (NAV) forecasting model was presented in [21] in which GA and PSO were implemented for ensemble weight optimization. In this work we propose an ensemble based automated medical image diagnosis system for diagnosing the healthy and diseased brain conditions by analyzing the brain MR images. Two-dimensional DWT (2D DWT) is used for feature extraction from the MRIs in the first step. PCA is applied for dimensionality reduction on the extracted wavelet coefficients in the next step. To get better classification performance we have implemented a novel ensemble based model for classification of the brain images. In this ensemble approach four different individual base classifiers namely ANN, KNN, SVM and Naïve Bayes are first employed separately for training. After successful validation of each model the output obtained by each individual model is optimally weighted using an improved firefly algorithm (improved FA) to produce the best possible results for the ensemble model as a whole. The layout of this paper is as follows. The pre-processing methods adopted for the image datasets are described in section II. The proposed ensemble classifier model is discussed in section III. Section IV gives the detail description of the proposed improved firefly algorithm and its implementation. The experimentations conducted, results comparison and validation are discussed in section V. The conclusion of this work is given in section VI.

II. PREPROCESSING THE IMAGE DATASETS

A. Feature Extraction using DWT

DWT is one of the effective tool that is widely used for feature extraction in case of brain MR image datasets. Through wavelets the different frequencies of an image can be effectively analysed at different levels of resolution. In this work 2D-DWT was used to decompose the MR images to get the wavelet coefficients that would work as the feature vectors for classification at the next level. It uses cascaded filter banks of both high-and low pass filters to decompose the signal into different sub-bands as shown in fig.1.

![Fig. 1. Schematic representation of 2-Dimensional DWT](image)

In 2D-DWT, image is processed along the two dimensions separately at each level by using low and high pass filters to generate four sub band images (LL, LH, HH, HL) at each scale. The lower frequency sub-band LL is considered as the approximation component resulting finer frequency resolution, whereas the other three sub-bands are the detailed components of the image. The LL sub-band is used for the next level of transformation as illustrated in fig.2. Thus DWT offer a simple hierarchical structure for inferring the image information. In this work we have used 3-levels decomposition of Haar wavelet to extract features for each brain MR image.

![Fig. 2. 3-level decomposition in 2D DWT](image)

B. Feature Reduction using PCA

PCA is one of the most widely used tool for dimensionality reduction in case of medical image data sets. PCA exploits the correlation among the features in the original dataset and performs an orthogonal transformation on the features to reduce them to a smaller set of linearly uncorrelated variables called principal components. The principal components represent the reduced set of features that retains most of the variations in the original data and are determined by finding the largest eigen vector of the correlational matrix. PCA results in an ordered list of principal components such that the first component is the maximum contributor to possible variance and the successive components in the list contributing less than its previous one figures and tables.

III. PROPOSED ENSEMBLE CLASSIFIER

In the field of medical data analysis, even a slightest improvement in the accuracy of the classification technique is of significance. Ensemble learning methods are now widely used for classification tasks and are preferred over single classification model [13]. Ensemble models offer higher generalization ability compared to the individual base models by eliminating the bias associated with the base models. In ensemble technique multiple classification models are first trained with the dataset and then their results are combined in some way to produce more accurate and improved results. The constituent models of an ensemble model are called the base models. Ensemble models assure better performance when there exist diversity among the base models [22]. In this work we developed an ensemble based classification model for diagnosing the MR brain image as normal or diseased. The ensemble model combines the results of four heterogeneous classification methods as a post-processing task. The proposed classification model includes two steps. In the first stage four base classifier models, namely ANN, KNN, Naïve Bayes and SVM are employed for training on the image dataset. The base classifiers are chosen so as to offer linear mapping (SVM), non-linear mapping (ANN) and probabilistic approach (k-NN) and there by caters to the diversity need of ensemble model. In the second step towards designing the ensemble model, an improved firefly optimization algorithm (improved FA) is used to optimally decide the weightage or contribution of the four base models to the overall ensemble model by considering the mean square error (MSE) of the overall ensemble model as the fitness function.
Here weights assigned to each base classifier are considered as free parameters and their combination values are optimized using the proposed improved FA to get the best classification performance.

Thus the linear combination of the four base classifiers representing the weighted sum of the individual classification models can be written as
\[ y = w_1y_1 + w_2y_2 + w_3y_3 + w_4y_4 = \sum_{i=1}^{4} w_i y_i \] (1)
Where, \( y \) is the final predicted output of the ensemble system, \( y_i \) are the individual output and weight associated with the base classifier models such that \( w_i (i=1,2,\ldots,4) > 0 \) and \( w_1 + w_2 + w_3 + w_4 = 1 \)
This proposed ensemble model is shown in figure

Fig. 3. Proposed ensemble classifier model.

IV. IMPROVED FIREFLY ALGORITHM

Firefly Algorithm (FA), proposed by Yang et al [23], is a swarm based metaheuristic optimization technique that is inspired by the light emitting nature of fireflies at night. The light flashing by the fireflies basically serves two important purposes, that is, to attract mating partners and to search for preys. The FA is formulated on the following three assumptions in relation to the movement of the fireflies

I: Regardless of their gender, fire flies get attracted and move towards brighter fireflies.

II: The brightness of the firefly at a particular position is determined by the value of the fitness function of the optimization problem. Higher the value of the fitness function of a firefly at a given position, the brighter the firefly is.

III: The attractiveness is directly proportional to brightness and brightness is inversely proportional to distance between a pair of firefly within the solution space.

When fireflies move toward brighter fireflies, their position and fitness values gets upgraded and so does their brightness and attractiveness. Hence every time fireflies update their position they tend to become even brighter and since brightness of a firefly represents its fitness, the swarm of fireflies eventually tends to move to a better position in the solution space with time.

Thus the FA implementation is based on two major elements, namely brightness and attractiveness. Brightness decides the location superiority of a firefly whereas attractiveness decide the movement of the fireflies

Attractiveness of a firefly at position \( r \) is given by (2).
\[ \beta (r) = \beta_0 e^{-\gamma r^2 y} \] (2)
Where \( \beta_0 \) represents the intensity of the firefly at \( r = 0 \) and \( y \) is the fixed light absorption coefficient at source.

The Euclidean distance between two fireflies \( x_i \) and \( x_j \) is calculated as
\[ r_{ij} = \| x_i - x_j \| = \left( \sum_{k=1}^{d} (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}} \] (3)

The movement of firefly \( i \) attracted to another more attractive firefly \( j \) is given by (4)
\[ x_i = x_i + \beta_0 e^{-\gamma r^2 y} (x_j - x_i) + \alpha (rand - 0.5) \] (4)
Where, \( r_{ij} \) is the distance between \( x_i \) and \( x_j \), \( \beta_0 \) represents the attractiveness at \( r_{ij}, y \) is the fixed light absorption coefficient such that \( y \in [0,\infty) \), \( \alpha \) is the randomization parameter. \( rand \) is a random number generator between 0 and 1.

Though FA has proven its superiority in solving different optimization problems in engineering and other fields, like any other evolutionary algorithms, it also suffers from the problem of getting stagnant in local optima that does not corresponds to the global best solution [24]. An improved FA is proposed here that introduces a Mixed-Cauchy mutation into FA in order to evade premature convergence in FA by means of long jumps made by the Mixed-Cauchy mutation. The objective behind using Cauchy mutation in FA is to bring diversity to the population of fireflies in each generation and thereby enhance global search ability. Incorporating Cauchy mutation to the firefly movement equation help them get out of the local search space by making them to jump to a better position in the search space. As because in Cauchy distribution there exist no expectation and variance for the same is infinite, so Cauchy mutation allows a firefly to have a long jump. Now the mixed MixedCauchy – pdf used in this work can be generated using (5).

\[ MixedCauchy – pdf = \alpha Cauchy – pdf1 + (1 - \alpha) Cauchy – pdf2 \] (5)
Here \( \alpha \) is a random number between 0 and 1.

The equation for \( Cauchy – pdf1 \) and \( Cauchy – pdf2 \) are given in (6) and (7).
\[ Cauchy – pdf1 = \frac{1}{\pi y \left[ 1 + \left( \frac{x-y}{\gamma y} \right)^2 \right]^\frac{3}{2}} \] (6)
\[ Cauchy – pdf2 = \frac{1}{\pi y \left[ 1 + \left( \frac{x-y}{\gamma y} \right)^2 \right]^\frac{1}{2}} \] (7)

Where \( y_1 \) and \( y_2 \) represents the location parameters and \( x \) in our case represents the current best position of the firefly.

A. Improved Firefly Algorithm

Algorithm: Improved FA

1. **Objective function:** \( f(x) \).
2. **Generate initial Population:** \( N \).
3. **Max. Generations:** \( G \).
4. **Calculate attractiveness and light intensity I for each firefly \( x_i \in x \) in \( N \).
5. **while** (\( T < G \)) **do**
6. **for** \( i = 1 : N \) **do**
7. **for** \( j = 1 : i \) **do**
8. **if** \( (f_j > f_i) \) **then**
9. **Move firefly \( i \) to \( j \)**
10. **end if**
11. **Update \( \beta \) using**
12. **Evaluate new solution using**
13. **end for**
14. **end for**
15. Rank the fireflies in the order of their fitness and update the current global best position.
16. Apply Mixed-Cauchy mutation on the global best position and compare its fitness with the existing current global best to find the global best position.
17. \( T = T + 1 \)
18. **end while**
19. Return the global best position

**B. Procedure for improved FA based weighted linear ensemble classifier model**

The proposed improved FA optimized ensemble classifier model follows the following steps.
1. Each input image sample is processed through Individual base classifier models i.e. ANN, KNN, Naïve Bayes and SVM model to generate four different outputs.
2. The four generated outputs of the base models are associated with four different weights and then linearly combined to produce the final predicted output.
3. This predicted output is compared with the actual output to produce the error (MSE is calculated).
4. Improved FA is employed to minimize the MSE.
5. Following each iteration, the MSE values are plotted to generate the convergence characteristics.
6. When the convergence curve attains the lowest optimum (MSE reaches the optimum lowest value) training is stopped and the best firefly position refers to the optimal weights.
7. Using these optimal weights, the test datasets are taken for performance comparison.

**V. EXPERIMENTATION, RESULT ANALYSIS AND VALIDATION**

All experiments were conducted using MATLAB (R2016) on a personal computer with 3.30 GHz Core-i5 processor having 4GB RAM running under Windows 10 operating system. The proposed hybrid classifier model is experimented on three brain image datasets collected from [25] supported and maintained by Medical School of Harvard University. These datasets hold T2-weighted brain MR images in gif format with size 256×256 in transaxial plane. For our purpose we collected three different brain MR image datasets for three different brain diseases, namely, Cavernous Angioma, Chronic Subdural Hematoma, Vascular Dementia having. The number of MRI samples of both diseased and normal brain in the above three collected datasets were 73, 93 and 91 respectively.

Process of feature extraction and dimensionality reduction plays an important role in deciding the classification performance. In this work 2D-DWT was used to extract the key features from the brain MRIs. 3-levels decomposition via Haar wavelet was used to extract 1024 features for each brain MR image. In the next step PCA was applied for feature reduction on the finer features extracted using wavelets. As per the standard procedure the input feature vectors were normalized before applying PCA. The threshold for variance taken for selecting the principal components is 95%. Figure No. 4 shows the variance with respect to the first 60 principal components for the three datasets. The dimensionality of the wavelet coefficients were thus reduced from 1024 to 24, 15, 12 for Cavernous Angioma, Chronic Subdural Hematoma and Vascular Dementia dataset respectively.

Four base classifiers, namely ANN, SVM, KNN and Naive Bayes were independently trained with respect to their objective functions by taking the input image samples from the training set of the three datasets. Hold-out method was used for training and testing.
The final parameters for each of the classifier model were frozen after the training was completed. Then the individual outputs of each of the model were combined through the weighted linear ensemble model as per (1). The ensemble weights were optimized using the proposed improved FA. After completion of the training process, the optimal weight values obtained were taken for the testing phase. The parameters and their values considered for optimization are given below Table No. 1

### Table No. 1. Improved FA implementation parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of decision variables</td>
<td>4</td>
</tr>
<tr>
<td>Maximum number of iterations</td>
<td>100</td>
</tr>
<tr>
<td>Number of fireflies</td>
<td>50</td>
</tr>
<tr>
<td>Light absorption coefficient ($\gamma$)</td>
<td>1</td>
</tr>
<tr>
<td>Attraction coefficient ($\beta_0$)</td>
<td>2</td>
</tr>
<tr>
<td>Mutation coefficient ($\alpha$)</td>
<td>0.2</td>
</tr>
<tr>
<td>Location parameters ($y_1$, $y_2$)</td>
<td>2, 4</td>
</tr>
</tbody>
</table>

To compare and validate the performance of our proposed method, experiments were also conducted to optimize the ensemble weights using another three established bio-inspired optimization techniques i.e. FA, PSO and GA. Classification accuracy is considered to measure the performance of these models. The classification accuracy is calculated as:

\[
\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples classified}} \times 100\%
\]

The accuracy obtained for all experimented models are given in Table No. 2. Our improved FA base ensemble model achieved 96%, 97.8% and 98.9% of classification accuracy in case of Cavernous Angioma, Chronic Subdural Hematoma and Vascular Dementia dataset respectively. The results show that all the four ensemble based models perform better than the individual models applied for classification. At the same time it is also evident that the proposed improved FA based ensemble model outperforms all other models and achieves highest classification accuracy in all the three datasets.

### Table No. 2. Classification Accuracy (in %) achieved for all methods in the three datasets.

<table>
<thead>
<tr>
<th>Method/Model</th>
<th>Cavernous Angioma</th>
<th>Chronic Subdural Hematoma</th>
<th>Vascular Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>85.5</td>
<td>83.8</td>
<td>78.0</td>
</tr>
<tr>
<td>KNN</td>
<td>86.8</td>
<td>81.7</td>
<td>82.4</td>
</tr>
<tr>
<td>SVM</td>
<td>89.4</td>
<td>89.5</td>
<td>85.7</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>88.4</td>
<td>87.1</td>
<td>84.7</td>
</tr>
<tr>
<td>Ensemble-GA</td>
<td>89.4</td>
<td>93.5</td>
<td>85.7</td>
</tr>
<tr>
<td>Ensemble-PSO</td>
<td>92.1</td>
<td>94.6</td>
<td>91.9</td>
</tr>
<tr>
<td>Ensemble-FA</td>
<td>93.4</td>
<td>95.7</td>
<td>92.3</td>
</tr>
<tr>
<td>Ensemble-Improved FA</td>
<td><strong>96.0</strong></td>
<td><strong>97.8</strong></td>
<td><strong>98.9</strong></td>
</tr>
</tbody>
</table>

Thus from the experimental study it is evident that the improved FA based ensemble model outperforms the individual classifier models as well as the other bio-inspired optimization based ensemble classifier models.

The MSE values were plotted at each of the 100 iterations. The comparison of the error convergence curves with other nature inspired algorithms i.e. FA, GA and PSO for all three datasets are given in Figure 4. It is observed that improved FA achieves minimum MSE values with a better convergence rate compared to the other three.
VI. CONCLUSION

In this study we proposed and validated a weight optimized multiple classifier based ensemble model for detection of normal and pathological brain condition from MR brain images. The improved FA optimized model found to perform significantly better compared to all the individual and other ensemble models studied in this work. Thus our proposed model may be used to develop computer aided diagnosis (CAD) systems for more accurate diagnosis of different brain diseases from MR images. This study can be extended to test the performance of the proposed model on multi-class classification of brain image datasets.

REFERENCES


AUTHORS PROFILE

Debahuti Mishra: She is working as Professor and Head of the Computer Science & Engineering Department of Siksha O AmusandhanDeemed to be University. His research area includes medical image analysis, machine learning and soft computing.

Sarada Prasanna Pati: He is an Assistant Professor and pursuing Ph.D. in Computer Science & Engineering Department of Siksha O AmusandhanDeemed to be University. His research area includes Data Mining and soft computing techniques.