Abstract—The paper proposes a novel approach to recognize smell stimuli from the electroencephalogram (EEG) signals acquired during the period of inhalation. The main contribution of the paper lies in feature selection by an evolutionary algorithm and pattern classification by Differential Evolution induced Hopfield neural network. One additional merit of the work lies in data point reduction by Principal component analysis. Experiments undertaken on 25 subjects with 10 smell stimuli indicate that the proposed scheme of feature selection, data point reduction and classification outperforms the traditional approach by a wide margin. Experimental results confirm that the smell stimuli excites the pre frontal lobe of the human brain and is responsible for a special type of brain rhythms (EEG signal) in alpha-band, theta-band and delta-band.

Keywords—olfaction; electroencephalogram signal; differential evolution; principal component analysis; Hopfield neural network.

I. INTRODUCTION

Olfaction refers to the sense of smell involved in identification of foods, fragrances, and chemical stimuli. This paper attempts to examine the effect of olfaction on electroencephalogram (EEG) [1] signals. Apparently, odor encoding and perception involves memory cells, namely long-term memory, to retrieve and match odor of known stimuli [2]. Participation of neurons in smell stimuli processing vary widely depending on the nature and complexity of odor detection. Here, we analyze the EEG pattern during different odorant inhalation followed by the theoretical guidelines and distinguish the different odorants with the basis of the extracted features regarding the stimuli.

Signal modality is an important issue in EEG research. In the present context of olfaction recognition, we used SCP as the modality of EEG signal. Slow cortical potential, as found in previous study [3], may be defined as gradual changes of potential spotted in the membrane of dendrites that lasts from 300ms up to seconds. SCP originates due to feedback and positive reinforcement mechanism [3]. Here, the positive potential shift of SCP is generated during smell recognition state.

Selection of classifier has been guided by two issues here: i) fewer number of classes and ii) possibility of misclassification by the source noise in smell stimuli. Hopfield neural net [4] here is an ideal choice for the aforesaid reasons. First, the energy function of the Hopfield net can be tailored to have fewer minima corresponding to the stable state of the Hopfield dynamics. Second, noisy stimuli ultimately settle down to the desired optima as the energy surface contains fewer and widely spaced optima. In this paper, we optimally select the weight matrix by minimization of the energy function of the Hopfield dynamics using Differential Evolution (DE) algorithm [5].

The evolutionary feature selection algorithm constructed for the present application aims at identifying the essential features of the data points so as to satisfy two criteria jointly. First, each selected feature should have a narrow valuation space for all the data points lying under each class. Second, for each selected feature i, the class mean of the feature for any two classes should differ by a wide margin. An objective function is created to jointly optimize the above two objectives, and a DE algorithm has been employed to determine the essential features that jointly optimize the given objective function.

Very few works on EEG analysis for smell stimuli classification are found in the literature [6]–[9]. One of the pioneering works in this regard is due to Boeijinga and Lopes da Silva [6]. They examined the possible correlation between olfactory stimulus and the nonlinearity in EEG. Inspired this work, Harada et al. [7] attempted to identify possible correlation between olfactory stimulus and brain functionality induced by EEG. Almost at the same time, Nicolas et al. [8] studied the effect of odor at several consumers’ behavior. In [9] also, EEG signals are extracted to classify personal preferences of olfactory stimulus by using factor analysis. Henkin and Levy [7] examined the correlation of pleasant (unpleasant) stimuli with EEG signal taken from left (right) hemisphere. In fact, by applying theories and computational intelligence, researchers developed a technical device, known as ‘electronic nose’ [10], which can successfully detect some significant non-odorables gases. Unfortunately, none of the above works directly dealt with classification of source of odors from EEG signals. This paper proposes a novel approach to classify 10 different sources of odors using EEG signals with special emphasis in i) exclusive feature selection, ii) data point reduction and iii) classification of smell stimuli.
The paper is divided into five sections. Section II provides principles and methodology of the proposed scheme. Section III presents experimental frameworks and results. Section IV gives performance analysis of the proposed classifier model. Finally, conclusions of our findings are listed in Section V.

II. PRINCIPLES AND METHODOLOGY

This section provides an overview of the proposed scheme to determine the class of an unknown smell. Fig. 1 illustrates the basic steps of the smell recognition process.

![Fig. 1. An Overview of the proposed method](image)

A. Feature Extraction

Feature extraction is an important tool that extracts various kinds of well-known features such as time domain features, frequency domain features and time-frequency domain features from EEG signal. Here, we consider Hjorth parameters [11], Autoregressive (AR) parameters [12], Power Spectral Density (PSD) [13] and Wavelet coefficients [14]. The dimension of the extracted features varies from 3 to 1878.

B. Feature Selection

This section provides an interesting approach to feature selection. Let, we have a set of N, D-dimensional data points $X = \{ \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_N \}$, where, $\tilde{X}_j = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}]$ are the $j$-th feature vector of dimension $D$ for the $i$-th pattern belonging to the class $j$. $\mu_j$ be the mean of the $j$-th feature in class $j$. $\sigma_j^l$ be the standard deviation of the $j$-th feature in class $l$. $N_j$ be the data points in class $j$.

Now, the optimization problem using Differential Evolution (DE) to be developed to handle the problem is built up using the following two strategies.

1. Individual feature values of all the data points within a class should have small difference. This is ensured by minimizing (1).

$$J_1 = \sum_{j=1}^{K} \sum_{i=1}^{N_j} \sum_{j=1}^{N_j} \sum_{k \neq j}^{N_j} \left| X_{i,j} - X_{i,k} \right|$$

2. The mean/standard deviation of a feature in any two classes should be maximized. This is ensured by maximization of (2).

$$J_2 = \sum_{j=1}^{K} \sum_{m=1}^{d} \sum_{i=1}^{N_j} \frac{\mu_j - \mu_m}{\sigma_j}$$

A composite objective function is constructed by minimization of (3).

$$J = J_1 - \lambda J_2$$

Here, $\lambda$ is the Lagrange’s multiplier used to scale $J_1$, so that $J_2$ and $\lambda J_2$ have the same order of magnitude. Solving the optimization problem requires employing DE to optimize the objective function (3) with an aim to determine an optimal selection of features $i$ in $[1, K]$. The desired encoded value of lies in $[1, 2^d - 1]$.

Pseudo code for feature selection using DE

Input: $D$ dimensional data points $X = \{ \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_N \}$, where $\tilde{X}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}]$, each having $D$ features and an assigned class level $l \in [1, K]$ for $K$ classes with class labels for each $\tilde{X}_i$.

Output: Selected dimensions of the data points corresponding to minimal $J$.

Begin

1. Initialization: Initialize $NP$ number of trial solutions $\tilde{Z}_i$ of the format given in Fig.2 for $i = 1$ to $NP$. Initialize crossover ratio $Cr = 0.7$.

2. Mutation: For each $\tilde{Z}_i$, pick up 3 companion target vectors, $\tilde{Z}_j$, $\tilde{Z}_k$, and compute $\tilde{Z}_l = \tilde{Z}_i + F (\tilde{Z}_j - \tilde{Z}_k)$, where, $F$ is a scale factor in $[0, 2]$. Here $j$, $k$, and $l$ are mutually exclusive to each other and all are different from $i$.

3. Recombination: Now for each pair of $\tilde{Z}_i$ and $\tilde{Z}_l$, construct a new trial vector $\tilde{M}_i$, where $m$-th element of $\tilde{M}_i$ is obtained by $m_{i,j} \leftarrow z_{i,j}$, if $r$, a randomly selected number in [0, 1] < $Cr$. $m_{i,j} \leftarrow z_{l,j}$, otherwise.

4. Selection: For each pair of $\tilde{M}_i$ and $\tilde{Z}_i$, $\tilde{Z}_j \leftarrow \tilde{M}_i$, if $f(\tilde{M}_i) < f(\tilde{Z}_j)$, where $f(\cdot)$ is the fitness (objective) function.

5. Repeat from step 2 until the stopping criterion is satisfied.

End

C. Data Reduction

Here, we use Principal Component Analysis (PCA) [15], which has already been mentioned in our previous work [16] to find the class representatives of individual class.

Input: A set of $n$ feature vectors $FV$ of dimension $n \times d$, with each feature $FV_i$ of dimension $d < D$ (obtained after feature reduction) with $i \in [1, n]$ representing a particular smell class.

Output: The corresponding class representative $\hat{\theta}$ of dimension $d \times 1$.

Begin

1. Normalization: Let the $i$-th feature vector of dimension $1 \times d$ is designated by $FV_i = \{ f_{i,1}, f_{i,2}, \ldots, f_{i,d} \}$. To get the normalized $FV_i$, we use the following transformation.
Data Adjust = \mathbf{FV} - \mathbf{f} \bar{u} \tag{8}

Here \bar{u} is a vector of dimension 1 \times d with u_k = 1 for k = [1, d]. In other words, the Data Adjust can be defined as follows.

\[
\begin{bmatrix}
\mathbf{FV}_1' \\
\mathbf{FV}_2' \\
\vdots \\
\mathbf{FV}_n'
\end{bmatrix} = \begin{bmatrix}
f_{v_{1,1}}' & \cdots & f_{v_{1,d}}' \\
\vdots & \ddots & \vdots \\
f_{v_{n,1}}' & \cdots & f_{v_{n,d}}'
\end{bmatrix} \begin{bmatrix}
\mathbf{f}_{v_{1,1}} - \bar{f}_1 & \cdots & \mathbf{f}_{v_{1,d}} - \bar{f}_1 \\
\vdots & \ddots & \vdots \\
\mathbf{f}_{v_{n,1}} - \bar{f}_n & \cdots & \mathbf{f}_{v_{n,d}} - \bar{f}_n
\end{bmatrix}
\tag{9}
\]

where \mathbf{FV}_i' = [f_{v_{i,1}}', f_{v_{i,2}}', \ldots, f_{v_{i,d}}'] is the normalized mean adjusted i-th feature vector.

5. Evaluation of Co-variance Matrix: The co-variance matrix of Data Adjust is obtained from the outer product of matrix Data Adjust with itself. The matrix thus obtained is of dimension n \times n and is Co-variance matrix, symbolized by \mathbf{C}.

6. Evaluation of Eigen Values: The Eigen values of the matrix \mathbf{C} is obtained by determining the roots of the characteristic equation given by

\[|\mathbf{C} - \lambda \mathbf{I}| = 0\tag{12}\]

where \mathbf{I} is the identity matrix of dimension n \times n and | | denotes the determinant of the matrix. There would be n Eigen values of matrix \mathbf{C}.

7. Evaluation of Eigen Vectors: The i-th Eigen vector \mathbf{EV}_i corresponding to the i-th Eigen value \lambda_i for i = [1, n], is obtained by satisfying the following equation.

\[\mathbf{C} \mathbf{EV}_i = \lambda_i \mathbf{EV}_i\tag{13}\]

8. Principal Component Evaluation: By ordering the Eigen vectors in the descending order of Eigen values (largest first), one an ordered orthogonal basis is created with the first Eigen vector \mathbf{EV}_1 (corresponding to the highest Eigen value \lambda_{\text{large}}) having the direction of largest variance of the data. In fact, in this way, it turns out that \mathbf{EV}_1 is the Principal Component \mathbf{PC} (n \times 1) of the data set. Therefore,

\[\mathbf{PC} = \mathbf{EV}_1 = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}^T\tag{14}\]

corresponding to \lambda_{\text{large}} > \lambda_i, where i = [1, n].

9. Projection of Data Adjust along the Principal Component: Now, to get the class representative, the formula applied is given in (15).

\[\bar{\theta} = (\mathbf{PC} \times \text{Data Adjust})^T\tag{15} \]

End

Thus we get a class representative \bar{\theta} of dimension d \times 1 corresponding to n feature vectors of the particular class. This procedure is repeated for each of the K classes to obtain K class representatives.
D. Classification

This section presents a novel approach to classify smell stimuli from EEG signals obtained from the subjects during olfaction. The problem here can be realized by applying a classifier on a recurrent topology. It is important to mention here that usually unsupervised learning is used for clustering problem. However, here we attempt to realize classification problem using unsupervised learning technique. Fig. 2 gives the detailed view of classification process.

![Diagram showing the overview of classifier training using Differential Evolution](image)

**Fig. 2.** Overview of classifier training using Differential Evolution

1) Classifier Training

A Hopfield net is a recurrent topology, where the neurons have a given dynamics. Combining the dynamics of all the neurons, we ultimately obtain the system equation (16):

$$\mathbf{C} \frac{d\mathbf{u}}{dt} = -\mathbf{a}\mathbf{u} + \mathbf{W}\mathbf{x} + \mathbf{b}$$

where, $\mathbf{a} = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \alpha_d \end{bmatrix}$ is a diagonal matrix of $(d \times d)$;

$\mathbf{u} = [u_1, u_2, ..., u_d]^{T}$ is a vector of $(d \times 1)$, whose component $u_i$ denotes the system state of the $i$-th neuron;

$\mathbf{b} = [b_1, b_2, ..., b_d]^{T}$ is a minimization input vector of $(d \times 1)$, whose $i$-th component $b_i$ denotes the input of the $i$-th neuron;

$\mathbf{W} = [W_{ij}]_{d \times d}$ is a weight matrix of $(d \times d)$, whose $i$-th component denotes the connectivity strength from neuron $i$ to neuron $j$; we here presume that $W_{ij} = W_{ji}$ (symmetric weight).

Here $\mathbf{C} = \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & C_d \end{bmatrix}$ is a diagonal matrix, whose $(i,j)$-th component $C_i$ denotes the time constant of the neuron $i$.

For stability analysis of a Hopfield dynamics we usually construct an energy function if the form given in (17) is:

$$E(\mathbf{x}) = \frac{1}{2} \mathbf{x}^{T} \mathbf{C} \mathbf{W} \mathbf{x} - \mathbf{b}^{T} \mathbf{x}$$

Here, $\mathbf{x} = f(\mathbf{u}) = G\mathbf{u}$ (say),

and $\mathbf{G}$ is a $(d \times d)$ gain matrix.

Now, at equilibrium, we set $\frac{d\mathbf{u}}{dt} = 0$ in (16) yielding

$$\mathbf{a}\mathbf{u} = \mathbf{W}\mathbf{x} + \mathbf{b}$$

From (18) and (19), we have

$$\mathbf{u} = (\mathbf{aI} - \mathbf{W}\mathbf{G})^{-1} \mathbf{b}$$

Substituting (18) and (20) in (17) we obtain,

$$E(\mathbf{x}) = \frac{1}{2} (\mathbf{G}\mathbf{u})^{T} \mathbf{W}(\mathbf{G}\mathbf{u}) - (\mathbf{G}\mathbf{u})^{T} \mathbf{G}\mathbf{u}$$

$$= -\frac{1}{2} \mathbf{u}^{T} \mathbf{G}^{T} \mathbf{W} \mathbf{u} + \mathbf{b}^{T} \mathbf{G} \mathbf{u}$$

$$= -\frac{1}{2} (\mathbf{aI} - \mathbf{WG})^{-1} \mathbf{b}^{T} \mathbf{G}^{T} \mathbf{W} \mathbf{G} ((\mathbf{aI} - \mathbf{WG})^{-1} \mathbf{b})$$

$$= -\mathbf{b}^{T} \mathbf{G} ((\mathbf{aI} - \mathbf{WG})^{-1} \mathbf{b})$$

Now, for a given $\mathbf{b}$, minimization of $E$ with respect to $\mathbf{W}$ returns the optimal solution of $\mathbf{W}$. This is repeated for all class representatives given in (22).

$$\sum_{i=1}^{K} E(\tilde{b}_i) = \frac{1}{2} \left( (\mathbf{aI} - \mathbf{WG})^{-T} \mathbf{b} \right)^{T} \mathbf{G}^{T} \mathbf{W} \mathbf{G} \left( (\mathbf{aI} - \mathbf{WG})^{-T} \mathbf{b} \right)$$

### Pseudo-Code for the energy function optimization of Hopfield neural network using DE algorithm

**Input:** $K$ class representatives $\mathbf{b} = [\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_K]^{T}$, with each $\tilde{b}_i$ ($i \in [1, K]$) of dimension $d$, obtained after data reduction using Principal Component Analysis for each of the $K$ classes, a diagonal matrix $\mathbf{a}$, gain matrix $\mathbf{G}$, each of dimension $d \times d$.

**Output:** Optimal connection matrix $\mathbf{W}$ of dimension $d \times d$.

I. Set the generation number $i=0$ and randomly initialize a population of $NP$ individuals $P_t = \{\mathbf{W}_1(t), \mathbf{W}_2(t), ..., \mathbf{W}_{NP}(t)\}$ with $\mathbf{W}_i(t) = [W_{ij}(t)]$ (j, $k \in [1, d]$) of dimension $d \times d$ with $i \in [1, NP]$.

II. Evaluate the trial vector $\mathbf{W}_i(t)$ by measuring its cost function $f(\mathbf{W}_i(t))$ by (10).

III. $\mathbf{W}_{best}(t) \leftarrow \arg \min_{i \in [1, NP]} (f(\mathbf{W}_i(t)))$

IV. While the stopping criterion is not reached do

    Begin
    For $i=1$ to $NP$
    Begin
    End
    End

Begin
IV a. Mutation:
Generate a donor vector
\[ V_i(t) = [V_{i,j,k}(t)] \] corresponding to the \( i \)-th target vector
\[ W_i(G) \]
via the following mutation scheme of DE.
\[ V_i(t) = W_{\text{rand1}}(t) + F(W_{\text{rand2}}(t) - W_{\text{rand3}}(t)) \]
where \( \text{rand1}, \text{rand2} \) and \( \text{rand3} \) are mutually exclusive integers randomly chosen from the range \([1,N]\), and all are different from the base index \( i \). \( F \) is the scale factor.

IV b. Crossover:
Generate trial vector \( U_i(t) = [U_{i,j,k}(t)] \) for the \( i \)-th target vector \( W_i(G) \) through binomial crossover as described below.
\[ U_{i,j,k}(t) = \begin{cases} V_{i,j,k}(t) & \text{if } \text{rand}_{j,k} \leq \text{CR} \\ W_{i,j,k}(t) & \text{otherwise} \end{cases} \]
for \( j, k = [1, d] \) and \( \text{CR} \in [0, 1] \) is the cross-over rate. Here \( \text{rand}_{j,k} \) is a uniformly distributed random number lying in \([0, 1]\) and is initiated independently for each \( j, k \)-th component of the \( i \)-th vector.

IV c. Selection:
Evaluate the trial vector \( U_i(t) \) by measuring its cost function \( f(U_i(t)) \).
\[ \text{IF } f(U_i(t)) < f(W_i(t)) \]
\[ \text{THEN do} \]
\[ \text{Begin} \]
\[ W_i(t+1) = U_i(t); \]
\[ \text{IF } f(U_i(t)) < f(W_{\text{best}}(t)) \]
\[ \text{THEN do} \]
\[ \text{BEGIN} \]
\[ W_{\text{best}}(t) = U_i(t); \]
Evaluate \( f(W_{\text{best}}(t)) \) and save it for future.
\[ \text{End If;} \]
\[ \text{End}; \]
\[ \text{Else } W_i(t+1) = W_i(t); \]
\[ \text{End If;} \]
\[ \text{End For;} \]
\[ \text{IF } d \text{. Increase the counter value } G = G+1. \]
\[ \text{End While;} \]

2) Classifier Testing
The classification process of an unknown feature vector \( P \) (of dimension \( d \times 1 \)) by the continuous Hopfield network is outlined below (Fig.3), where \( \tilde{\theta} \) is the input class representatives derived from PCA, and \( \tilde{X} \) is the output smell class.

![Fig. 3. Block diagram of Hopfield net testing](image)

1. Obtain the information stored in the Hopfield neural network with connectivity matrix \( W \) corresponding to the \( k \)-th class representative \( \tilde{\theta}_k \) denoted as \( \tilde{u}_k \), for \( k = [1, K] \) by calling \( \text{stored_info_HNN} \left( \tilde{\theta}_k \right) \).

2. Obtain the information stored in the Hopfield neural network with connectivity matrix \( W \) corresponding to the unknown feature \( \tilde{P} \) represented as \( u \), for \( k = [1, K] \) by calling \( \text{stored_info_HNN} \left( \tilde{P} \right) \).

3. The unknown vector is assigned the class level \( c \)
\[ d \left( \tilde{u}, \tilde{u}_k \right) = \min_{\forall k \in [1,K]} d \left( \tilde{u}, \tilde{u}_k \right). \]

Here, \( d \left( \tilde{u}, \tilde{u}_k \right) = \| \tilde{u} - \tilde{u}_k \| = \sqrt{\sum_{i=1}^{d} (u_i - u_{ki})^2} \).

Procedure \( \text{stored_info_HNN} \left( P \right) \)
1. Initialize \( \tilde{u} = 0 \).
2. Compute \( \Delta \tilde{u} = C^{-1} \tilde{P} \Delta t \) with \( \Delta t = 1 \) unit (say).
3. Set \( \tilde{u} = \tilde{u} + \Delta \tilde{u} \) where \( \Delta \tilde{u} = C^{-1}(-a \tilde{u} + Wx + \tilde{P}) \Delta t \) and \( \tilde{x} = f(\tilde{u}) = G \tilde{u} \).
4. Repeat step 3 until \( |\Delta u_i| \leq \delta, \forall i \in [1,d] \), where \( \delta \) is a pre-assigned positive number, however small.
5. Return \( \tilde{u} \).

III. EXPERIMENTS AND RESULTS

A. Experimental Paradigm
Fig. 4 shows the experimental set-up, where subjects are asked to comfortably sit on a relaxing chair with armrests with closed eyes. 10 different odorants including naphthalene, odonil, medicated talcum powder, cinnamon, rosewater, male perfume, hydrogen sulphite, ammonia, methane and camphor are placed in ten covered bottles of similar sizes. Each of 25 participants are given each odorant for 10 seconds and three minutes break between two different odorants to eliminate the effect of previous smell stimuli on the current EEG responses. The process is repeated for 10 times for each subject. Consequently, we have as many as 250 instances of EEG signals per stimulus.

![Fig. 4. Experimental Set-up](image)

B. Channel Selection
EEG data are continuously recorded from Fp1, Fp2, F3, F4, F7, F8 and Fz of a 21-channel Neurowin system according to the international 10-20 electrode placement method [17] (Fig.5). A1 and A2 electrodes are taken for reference. From the experiment, it has been found that Fp1 and Fp2 have greater influence of smell stimulus when EEG data from all
the electrodes are taken into consideration for frequency band calculation.

Fig. 5. Electrode positions for experiment according to 10-20 electrode placement system.

C. Frequency Band Selection

Decoding of EEG signals require the judicious selection of useful bands of frequency. In this paper, we extract alpha (8-13Hz), beta (13-30Hz), theta (4-8Hz) and delta (0.5-4Hz) band of frequencies to observe the change in band power (in dB). From the experimental data, it is observed that alpha, beta and delta band power significantly rises with time, whereas, beta band power has pure oscillatory changes in presence of stimulus. The observation here, validates the previous studies [18], reporting the fluctuations of alpha and theta bands for various smell stimuli. Based on the analysis, we design the band pass filter having pass band frequencies of 0.5-13Hz to remove noise and other artifacts.

D. Feature Selection

We extracted features for each one of 250 instances per stimulus and use principal PCA to reduce the 250 data points into a single one. A particular feature, when applied on several odorants, may produce different feature values for different smell stimuli. Figure 8 shows the variation in feature values of different odorant for a particular feature. From the figure, it is evident that odonil is best classified by considering hjorth parameters and PSD. It is important to note here that, there is a chance of misclassification between odonil and naphthalene if we only consider AR parameters. Similarly, in case of wavelet detail coefficient, feature values of all odorants lie almost in the similar ranges. To deal with the problem, we form three distinctive sets of features having following dimensions.

1) Set 1: Hjorth Parameters (dimension:3) and PSD (dimension:513) = Total number of features:516.
2) Set 2: Wavelet (dimension:1878) and PSD (dimension:513) = Total number of features:2391.
3) Set 3: Autoregressive Parameters (dimension:11) and PSD (dimension:513) = Total number of features:524.

E. Reduced Feature Selection

We apply DE-based evolutionary technique to select best possible features from the original feature sets with an ultimate aim to optimize classifier performance. DE selects 47, 189 and 59 best features from the feature sets containing respectively 516, 2391 and 524 features.

F. Classifier Performance

In this section, we compare the performance of our DE-Hopfield-NN algorithm with six standard classifiers including 1) PCA-LDA [19], 2) PCA-KNN [20], 3) PCA-FFNN [21], 4) PCA-LSVM [22], 5) PCA-SVM-RBF [23] kernel, and 6) PCA-Naïve Bayes [24]. In Table-I, we present the mean and the standard deviation (within parenthesis) of the classification accuracy obtained for 20 independent runs of each of the algorithms without feature selection. Of 250 trials, classifier is trained by 10-fold cross-validation of 1250 trials, and tested by remaining 1250 data to produce classification accuracy. Table-II reports the number of selected features, the mean and standard deviation of the final classification accuracy over the runs of different algorithms using the feature selection methodology. The best solution in each case has been shown in bold.

Here, we use paired t-tests to compare the means of the results produced by best and the second best algorithms. The t-tests are quite popular among researchers and they are fairly robust to violations of a Gaussian distribution with large number of samples like 25 [25]. In the 10th column of Tables I and II, we report the statistical significance level (here, ‘+’ indicates the t value of 49 degrees of freedom is significant at a 0.05 level of significance by two-tailed test, ‘-‘ means the difference of means is not statistically significant and ‘NA’ stands for Not Applicable, covering cases for which two or more algorithms achieve the best accuracy results.

<table>
<thead>
<tr>
<th>Features</th>
<th>Dimension</th>
<th>Classifier</th>
<th>LDA</th>
<th>kNN</th>
<th>FFNN</th>
<th>LSVM</th>
<th>SVM-RBF</th>
<th>Naïve Bayes</th>
<th>Hopfield NN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Statistical Significance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Positive Rate (a) and False Negative Rate (b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jhorth + PSD</td>
<td>516 (3+513)</td>
<td></td>
<td>48.731 (0.086670)</td>
<td>65.709 (0.069827)</td>
<td>62.603 (0.059564)</td>
<td>70.632 (0.053972)</td>
<td>71.660 (0.041626)</td>
<td>73.953 (0.032640)</td>
<td>83.074 (0.015198)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelet + PSD</td>
<td>2391 (1878+513)</td>
<td></td>
<td>49.803 (0.090374)</td>
<td>66.874 (0.074398)</td>
<td>63.416 (0.063009)</td>
<td>73.659 (0.055288)</td>
<td>74.137 (0.047341)</td>
<td>79.141 (0.035589)</td>
<td>87.557 (0.016821)</td>
</tr>
<tr>
<td>AR + PSD</td>
<td>524 (11+513)</td>
<td></td>
<td>46.579 (0.094333)</td>
<td>61.676 (0.085320)</td>
<td>60.952 (0.064283)</td>
<td>69.208 (0.058775)</td>
<td>72.773 (0.048046)</td>
<td>76.794 (0.039904)</td>
<td>80.343 (0.020419)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.582</td>
<td>9.4253</td>
<td>5.9344</td>
<td>2.7603</td>
<td>2.3059</td>
<td>2.1384</td>
<td>0.19253</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE I: MEAN CLASSIFIER ACCURACY AND STANDARD DEVIATION (WITHIN PARENTHESIS) WITHOUT FEATURE SELECTION ALONG WITH FALSE POSITIVE RATE (a) AND FALSE NEGATIVE RATE (b)
TABLE II: MEAN CLASSIFIER ACCURACY AND STANDARD DEVIATION (WITHIN PARENTHESES) WITH FEATURE SELECTION USING DIFFERENTIAL EVOLUTION ALONG WITH FALSE POSITIVE RATE (α) AND FALSE NEGATIVE RATE (β)

<table>
<thead>
<tr>
<th>Features</th>
<th>Dimension</th>
<th>PCA-LDA</th>
<th>PCA-ANN</th>
<th>PCA-AKNN</th>
<th>PCA-FNN</th>
<th>PCA-SVM-RBF</th>
<th>PCA-LSVM</th>
<th>PCA-NaïveBayes</th>
<th>DE-HopfieldNN</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hjorth + PSD</td>
<td>47</td>
<td>73.643 (0.090457)</td>
<td>82.149 (0.079565)</td>
<td>73.088 (0.063432)</td>
<td>83.076 (0.049335)</td>
<td>83.845 (0.034264)</td>
<td>83.886 (0.025903)</td>
<td>95.595 (0.011400)</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>α</td>
<td>7.5529</td>
<td>7.0665</td>
<td>6.7527</td>
<td>5.5172</td>
<td>5.4292</td>
<td>1.7148</td>
<td>0.043256</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wavelet + PSD | 182 | 75.163 (0.077475) | 82.771 (0.066368) | 75.204 (0.06667) | 85.867 (0.046744) | 88.350 (0.032746) | 89.373 (0.017868) | 97.061 (0.011661) | + |
|              | α       | 10.313 | 8.6064 | 4.1327 | 3.5531 | 3.269 | 2.7289 | 0.44764 |
|              | β       | 14.524 | 8.7686 | 20.663 | 10.58 | 8.381 | 8.3481 | 2.4914 |

AR + PSD | 59 | 72.597 (0.087253) | 80.337 (0.073915) | 72.125 (0.061336) | 80.615 (0.051438) | 81.269 (0.039647) | 81.304 (0.032533) | 96.228 (0.011565) | + |
|          | α       | 9.5085 | 5.073 | 2.9345 | 2.5985 | 1.9017 | 1.4164 | 1.5162 |
|          | β       | 17.895 | 14.59 | 24.94 | 16.786 | 16.829 | 17.286 | 2.4558 |

IV. PERFORMANCE ANALYSIS

1) We compare DE-based energy minimization technique of Hopfield NN with other evolutionary algorithm-based energy minimization method with same vector representation, including Particle Swarm Optimization (PSO) [26] and real coded Genetic Algorithm (GA) [27]. To compare the relative speed of convergence and quality of solution (accuracy) of DE with other two optimization algorithms, the mean value of the objective function (mean best fitness) taken over 20 runs versus function evaluations (FEs) is plotted in Fig. 6. The observation from the figure illustrates that DE-based prediction of connection weights of Hopfield NN outperforms all algorithms in terms of quality of solution.

2) For determining the performance of two algorithms for detecting the proper classification of data points we have applied McNemar’s test. Let \( f_a \) and \( f_b \) be two classifiers (‘1’ and ‘0’) outputs obtained by algorithm A and B, when both the algorithms used a common input phylogenetic sequences R. We now define a null hypothesis:

\[
P_{R,R} [f_a(x) = f(x)] = P_{R,R} [f_b(x) = f(x)],
\]

where, \( f(x) \) be the experimentally induced sign to map any data point \( x \) on to specific sign classes \( K \), where \( f(x) \) is one of \( K=2 \) classes. Let \( n_{01} \) be the number of examples misclassified by \( f_a \) but not by \( f_b \) and \( n_{10} \) be the number of examples misclassified by \( f_b \) but not by \( f_a \). Then following (13), we define a statistic,

\[
Z = \left( \frac{n_{01} - n_{10}}{\sqrt{n_{01} + n_{10}}} \right)^2
\]

We evaluate \( Z \) which denotes the comparator statistic of misclassification between the Hopfield network-based classification algorithm (Algorithm: A) and any one of the competitor algorithms (Algorithm: B) for the \( j \)-th dataset for desired number of features (d) equal to 40. After considering DE-Hopfield NN as reference algorithm, Table III is evaluated to obtain \( Z_j \) and the hypothesis is rejected if \( Z_j > \chi^2_{1,0.95} = 3.841459 \), which indicates that the probability of the null hypothesis is correct only to a level of 5% and so, we reject it and clearly shows that DE-Hopfield NN outperforms all its competitors except PCA-SVM-RBF and PCA-Naïve-Bayes. Using Friedman Test, we have determined the performance of classifiers by their rank only. In this test, DE-Hopfield-NN outperforms all the algorithms by obtaining average rank of 1, where PCA-Naïve Bayes and PCA-SVM-RBF respectively obtain average rank of 2 and 3.

![Fig. 6. Relative performance of DE, PSO and real coded GA to infer connection weights of Hopfield NN](image-url)

TABLE III: STATISTICAL SIGNIFICANCE REFERENCE ALGORITHM: DE-HOPFIELD NN

<table>
<thead>
<tr>
<th>Classifier algorithm used for comparison using desired features d=50</th>
<th>Parameters used for McNemar Test</th>
<th>Zj</th>
<th>Comments on acceptance/rejection of hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA-LDA</td>
<td>92</td>
<td>175</td>
<td>25.1835</td>
</tr>
<tr>
<td>PCA-KNN</td>
<td>92</td>
<td>167</td>
<td>21.1428</td>
</tr>
<tr>
<td>PCA-FNN</td>
<td>91</td>
<td>150</td>
<td>13.9585</td>
</tr>
<tr>
<td>PCA-LSVM</td>
<td>90</td>
<td>128</td>
<td>6.2798</td>
</tr>
<tr>
<td>PCA-SVM-RBF</td>
<td>87</td>
<td>111</td>
<td>2.6717</td>
</tr>
<tr>
<td>PCA-Naïve-Bayes</td>
<td>85</td>
<td>102</td>
<td>1.3689</td>
</tr>
</tbody>
</table>
V. CONCLUSION

The paper introduces a novel approach to olfaction recognition using EEG analysis in Hopfield-DE synergism. Experimental instances for 10 different smell stimuli have been used to validate the proposed olfaction recognition scheme. The wavelet coefficients and PSD features have been found to give best results of classification accuracy irrespective of classifier. Further, the above two features when fed to the proposed Hopfield-DE classifier yields the best classification accuracy, indicating the best performance of the composite feature-classifier system. Statistical testing realized by McNemar’s test also supports the above statement of multiple patterns describing the smell stimuli.

ACKNOWLEDGMENT

Funding by University Grant Commission, India for the Project: University with Potential for Excellence Program in Cognitive Science (Phase II) granted to Jadavpur University, India is gratefully acknowledged.

REFERENCES