



# Recognition of Human Emotion using Radial Basis Function Neural Networks with Inverse Fisher Transformed Physiological Signals

Abdultaofeek Abayomi<sup>1\*</sup>, Oludayo O. Olugbara<sup>1</sup>, Delene Heukelman<sup>1</sup>

<sup>1</sup>ICT and Society Research Group, Department of Information Technology,  
Durban University of Technology, Durban, 4001, SOUTH AFRICA

\*Corresponding Author

DOI: <https://doi.org/10.30880/ijie.2021.13.06.001>

Received 23 November 2019; Accepted 4 October 2020; Available online 31 August 2021

**Abstract:** Emotion is a complex state of human mind influenced by body physiological changes and interdependent external events thus making an automatic recognition of emotional state a challenging task. A number of recognition methods have been applied in recent years to recognize human emotion. The motivation for this study is therefore to discover a combination of emotion features and recognition method that will produce the best result in building an efficient emotion recognizer in an affective system. We introduced a shifted tanh normalization scheme to realize the inverse Fisher transformation applied to the DEAP physiological dataset and consequently performed series of experiments using the Radial Basis Function Artificial Neural Networks (RBFANN). In our experiments, we have compared the performances of digital image based feature extraction techniques such as the Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP) and the Histogram of Images (HIM). These feature extraction techniques were utilized to extract discriminatory features from the multimodal DEAP dataset of physiological signals. Experimental results obtained indicate that the best recognition accuracy was achieved with the EEG modality data using the HIM features extraction technique and classification done along the dominance emotion dimension. The result is very remarkable when compared with existing results in the literature including deep learning studies that have utilized the DEAP corpus and also applicable to diverse fields of engineering studies.

**Keywords:** Neural network, human emotion, physiological signal, feature extraction, fisher transform

## 1. Introduction

The understanding of emotion expressed by people is a big challenge in affective computing, human computer interaction and social communication. Deployed solutions of human emotion recognition tasks could play an important role in various intelligent affective communication systems [1]. Emotion is also very useful in our understanding of human social behaviours and also richly embedded in human non-verbal communications. An important part of human to human communication is an expected change in the emotional state of a subject [2]. This could indicate emphasizing or clarifying spoken words, expressing agreements or disagreements, comprehending intentions as well as interacting with others and environments [3]. Human emotion is related to physiological signals and are influenced by physiological changes, external events and relationship with others, therefore representing a complex and dynamic state of human mind [1]. However, physiological signals originate from the peripheral and the autonomic central nervous system, which is one of the most complex systems of the human body. The nervous system is fundamental in human behaviours because it empowers human with the ability to perceive, understand and react to environmental events such

as emotion, desires and thoughts that are transmitted to the nervous system through the neuropeptides. In addition, human brain, which is an essential component of the central nervous system, generates brain waves, which contain electrical signals that are collected using electrodes attached to the scalp. These EEG waveforms from the brain represent the wave pictures of the electrical activity in the brain. They are physiological signals that inherently capture human emotional states and are obtained from the nervous systems using bio-sensors [4].

The peripheral nervous signals that have been collected in the literature include Galvanic Skin Response (GSR), Skin Temperature (TEMP), Electrocardiogram (ECG), Electroencephalogram (EEG), Electromyogram (EMG), Respiration (RESP), Blood Volume Pulse (BVP), Heart Rate (HR) and Heart Rate Variability (HRV) [5]. In the literature, emotion recognition using physiological signals has received relatively less attention in comparison to the audiovisual methods of facial expression and speech [6-9]. Meanwhile, the audiovisual methods have some intrinsic drawbacks, they are capable of being easily faked and the subject needs to be within a perimeter defined by the camera or must always listen to an audio signal. However, physiological signals evolve automatically and spontaneously. They are human reactions over which they have less controls and are less influenced by social, language and cultural differences [4, 6, 8, 10].

The modeling of physiological systems and signals helps in gaining understanding and generating methods of analysis which could be applied in diverse fields of engineering including biomedical engineering, biological engineering, healthcare engineering, robotic engineering as well as electrical and mechanical engineering etc. Consequently, the significant rise in the generation of physiological data, alongside the development of big data intelligence, has enabled the extraction of new insights from these massive physiological signals. These include bioelectrical signals such as EEG and ECG; biomagnetic signals such as magnetic resonance imaging and computed tomography; biochemical signals including pressure of oxygen and carbon dioxide in respiration; and bioacoustic signals such as speech and ultrasound. In addition, as a physiological signal, the EEG is one of the many biological signals that can be utilised to control wearable robotic devices according to the human motion intention in robotic engineering discipline. The various devices used in harvesting the diverse physiological signals utilised also have connection to engineering. Hence this current study is well suited to many application areas in engineering fields of studies.

In this study, our objectives include the acquisition of physiological data from the DEAP corpus and analyzed it by introducing a shifted tanh-based normalization scheme after which the inverse Fisher transformation algorithm was applied. Secondly, we desire to extract distinctive features from the transformed physiological data using the HOG, LBP and HIM techniques which are popular shape, texture and pixel intensity distribution descriptors respectively in the digital image processing domain. Thirdly, we crave to utilize the RBFANN pattern classifier to compare the performances of the three different feature descriptors in recognizing human emotional state along the arousal, valence, dominance and liking dimensions. The arousal scale measures the intensity and activation level of an emotion while the valence represents the pleasantness or otherwise of an emotion. On the other hand, the dominance scale measures an individual control over an emotion being experienced while liking scale is a subjective rating indicating the tastes of like or dislike of an emotion [4, 8, 11-13].

## 2. Related Works

In comparison with audiovisual emotion channels consisting of facial expressions, gestures and speech, relatively few research works have been conducted over the past one decade on emotion recognition using physiological signals [1, 7-9]. While appreciable, but varied recognition accuracy results were recorded in these relatively few works, feature extraction remains an open issue in affective computing research [8, 14]. There are also wide disparities in the number of emotions to be recognized, the number and types of bio-signals measured, dataset used and its quality, number of subjects sampled, emotion stimulus, modality considered, emotion models and pattern recognizer employed [4, 8, 14].

The subject-independent approach is the focus of many current studies [15-17] as a large number of subjects ensure reliability of results. Consequently, we followed this approach in this current study as it addressed the apparent limitations of the subject dependent approach such as its low degree of generalization, lack of inter-person variability in emotional feelings and inability of applications based on it to be readily deployed for practical usage [6, 14, 18]. However, the most critical challenge in developing subject-independent emotion recognition system is the identification of the most discriminatory features among the subjects which obviously impacts recognition results obtained.

It was observed that by using different methods [1, 3, 8, 18-22], various results have been recorded in the literature as the issue of feature extraction which is germane for performance attained in a pattern recognition system still remain open. In this study, we have utilized the digital image feature descriptors of HOG, LBP and HIM because of their strong performances in pattern recognition studies [23-31]. Therefore, we carried out feature engineering by applying the descriptors to the DEAP physiological dataset and compared our method, features and results obtained using the RBFANN pattern classifier with those in literature. The performances of the RBFANN pattern classifier was observed to determine whether favorable and reliable results could be obtained with the various configurations employed.

Soleymani et al. [3] utilized physiological signals including RESP amplitude, TEMP, GSR and ECG, harvested from experiment subjects induced with video clips to classify emotion into the arousal (calm, medium and excited) and valence (unpleasant, neutral and pleasant) states. Time domain statistical features, including average standard deviation of raw signals, as well as the absolute values of the first and second derivative were extracted and trained with an SVM classifier. Using a subject-independent approach, a classification accuracy of 46.2% was obtained for the arousal state while 45.5% was recorded for the valence state. The study was extended to EEG signals from where Power Spectral Density (PSD) and spectral power asymmetry features were extracted and trained using an SVM classifier. A classification accuracy of 52.4% and 57.0% was obtained for the three states of arousal and valence respectively. These results are considered low and might not meet the required level of accuracy expected in a real-time implementation.

In another research study using the DEAP dataset, Jun-Wen et al. [32] categorized ten groups of emotional pictures separated into five classes based on the valence and arousal dimensions. Facial EMG of 113 experiment subjects consisting of young and senior adults stimulated with these pictures was acquired for each of the mapped classes and 16 sets of features relating to the frequency, amplitude, variability and predictability of the EMG signals were extracted. An SVM classifier was applied on these features and a classification accuracy from 75.6 % upward was obtained for the five affective classes and the baseline for all the individuals.

A Deep Belief Network (DBN) based system was introduced by [33] to automatically extract features from 4 channels raw physiological data consisting of 2 EOG and 2 EMG channels respectively under an unsupervised scheme while building classifiers that predict human emotion along the arousal, valence and liking classes. The classification accuracies obtained were 60.9%, 51.2% and 68.4% respectively which compares favourably with the results achieved with the Naive Bayes classifier.

A related study reported by [34] adopted a two hidden-layer DBN architecture configured with visible and hidden nodes as 128-10-10 to classify EEG signals of the DEAP dataset along the binary label scheme for valence, arousal, dominance and liking respectively using unsupervised training and future learning. Classification experiments was done along individual subject as well as across all subjects. An SVM classifier was also applied on the power spectral density as well as the DBN features in order to compare the manually extracted features with the DBN features. Recognition accuracies of 58.2% (valence), 64.3% (arousal), 65.1% (dominance) and 66.3% (liking) was achieved with the PSD features across the all subjects while 58.4%, 64.2%, 65.8% and 66.9% respectively was recorded for the DBN features. There is no significant difference between the results of the two features as the possibility of learning affective features through a deep learning and manually generated features approaches were explored.

The DEAP physiological dataset was utilized by [35] by applying Empirical Mode Decomposition (EMD) method to extract first difference of time series, first difference of phase, and the normalized energy features from EEG signals which were decomposed into Intrinsic Mode Functions (IMFs) and classification was done along the arousal and valence classes using the SVM classifier. A classification accuracy of 71.99% and 69.10% respectively was achieved for the arousal and valence classes using 8 EEG channels of Fp1, Fp2, F7, F8, T7, T8, P7, and P8 while 72.10% and 74.10% classification accuracies for arousal and valence was achieved with 32 EEG channels. These performances are better than the results obtained by the same authors using other methods and features such as the fractal dimension, sample entropy and discrete wavelength transformations.

EEG signals' characteristics such as spatial, frequency domain and temporal were integrated by [36] and mapped to a two-dimensional images from which EEG Multidimensional Feature Images (MFI) were built to represent varied emotions in EEG signals. A deep learning approach named CLRNN involving hybridizing the Convolution Neural Networks (CNN) and Long Short-Term-Memory (LSTM) Recurrent Neural Networks (RNN) was consequently applied on the EEG MFI obtained from the DEAP dataset. With each subject, an average best emotion classification accuracy of 75.21% was achieved with a 2s time window frame as against the available 60s window size containing all the EEG data per sample. In addition, the results of other classification methods such as k-NN, random forest and support vector machine with the features and method proposed by the authors are below the 75.21% recorded.

However, a multiple-fusion-layer based ensemble classifier of stacked autoencoder (MESAE) for recognizing human emotions was proposed by [37] involving the identification of the deep structure based on physiological-data-driven approach. Stable feature representations of the physiological signals were obtained as the unwanted noise in the physiological signals' features were filtered by three hidden layers in each stacked autoencoder. The stacked autoencoder ensembles were achieved by using an additional deep model and the physiological features are divided into many subsets based on various feature extraction methodologies with each subset separately encoded by a stacked autoencoder. The derived SAE abstractions were merged based on the physiological modality to create six sets of encodings which subsequently served as input to a three-layer, adjacent-graph-based network for feature fusion whose features were used for human emotion recognition along the binary arousal and valence emotion states. Average classification accuracies of 77.19% and 76.17% was achieved for the arousal and valence state respectively using the MESAE scheme with deep classifier while the accuracies reached 84.18% and 83.04% with ensemble classification schemes.

In another study, the LSTM recurrent deep neural network was applied by [38] to the EEG physiological signals in the DEAP dataset. An average recognition accuracy of 85.65% was achieved for the arousal class while 85.45% and 87.99% average accuracies were recorded for the valence and liking classes respectively. Similarly, features were

extracted by [39] from EEG signals for affective state modeling using the Russell's circumplex model. The SVM and random forest classifiers were applied on the EEG features of statistical measures, band power as well as higher order crossing extracted from the DEAP dataset to classify human emotion into the valence and arousal classes. The highest classification accuracy obtained with the bandwaves spectral power density features by the SVM classifier was 69.2% and 88.4% for the bipartite scheme of arousal and valence classes respectively while the tripartite scheme recorded 59.5% and 55.9%. However, the random forest classifier recorded its best classification accuracy of 74.0% (arousal) and 88.4% (valence) along the bipartite labeling scheme (high/low) with the statistical bandwaves features and 63.1% (arousal) and 58.8% (valence) with the same features but for the tripartite labeling scheme (high/medium/low). Diverse lower results were recorded for each and combined bandwaves (Delta ( $\delta$ ), Theta ( $\theta$ ), Alpha ( $\alpha$ ) and Beta ( $\beta$ )) by both the SVM and random forest classifiers for the bipartite/tripartite labeling scheme and statistical bandwaves/spectral power density features.

Lastly, a framework to automatically search for the optimal subset of EEG features using Evolutionary Computation (EC) algorithms including the Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Genetic Algorithm (GA), Simulated Annealing (SA) and Differential Evolution (DE) was introduced by [40]. This is aimed at removing inefficiency and redundancy resulting from high-dimensionality introduced by combining all possible EEG features. The framework used frequency, time and time-frequency domain features of EEG signals from which some discriminatory features were selected using the EC algorithm and the probabilistic neural network pattern classifier was applied to classify human emotions into four classes. The DE algorithm yielded the best recognition accuracies of 96.97% and 67.47% for the MAHNOB and DEAP datasets respectively. Though the results obtained are promising, the challenge with this framework and method is its computation complexity as it takes about 80 hours to achieve convergence. This is not ideal for a real-time situation where efficient, prompt and accurate classification are required. We therefore explored a shallow machine learning approach using our proposed method of physiological data transformation and feature extraction to determine if better results above those reported, including by the deep learning approaches could be obtained.

### 3. Materials and Methods

#### 3.1 Experimental Dataset

The DEAP dataset was developed by [7] using video clips stimuli to elicit human emotions from 32 subjects (16 females) and their physiological data such as the EEG, EOG, EMG, GSR, RESP, BVP and TEMP were concurrently collected as they watched 40 one-minute extracts of music video clips. These clips are capable of eliciting the target or reported felt emotions of anger, contempt, disgust, elation, envy, fear, guilt, hope, interest, joy, pride, relief, sadness, satisfaction, shame and surprise. The physiological signals as well as frontal face videos of 22 subjects were acquired using various sensors and active electrodes with the Biosemi Active II system. However, only two modalities comprising the central nervous system, for instance the EEG, and the peripheral nervous system physiological signals consisting of the EOG, EMG, GSR, RESP, BVP and TEMP data as well as a fusion of these two modalities were specifically considered in our study. During each trial, subjects also undertook and reported a self-assessment of their degrees of valence, arousal, dominance and liking on a continuous 9-point scale using a Self-Assessment Manikin (SAM) while familiarity was rated on a 5-point scale. For each subject, there were 40 trials, which gives a total of 1280 samples for all the 32 subjects.

#### 3.2 Pre-processing of the Physiological Signals

In order to process the raw DEAP physiological data using our proposed methodology, firstly, we normalized the data channel by channel. There are different normalization schemes available in the literature amongst which the Min-Max and Z-score are famous. In particular, the Z-score has been used in the study of physiological data [41] because it has capability to dramatically simplify clinical interpretations [42]. However, both the Min-Max and Z-score normalization have received criticisms as they are both sensitive to outliers and their performances are sometimes poor [43]. The Min-Max normalization scheme also usually scale data to a fixed range of 0 to 1 thereby giving smaller standard deviations. But the Z-score normalization is often preferred to the Min-Max scheme especially when applied with Principal Component Analysis (PCA) procedure in order to compare similarities between features as the components that maximizes the variance is often the focus [44]. The tanh estimator therefore is suggested as a robust scheme in place of Min-Max and Z-score because of its efficiency and elegance [43] as a simple normalization scheme based on the tanh function is hereby introduced.

The general tanh function is given as;

$$t(u(x)) = \frac{e^{u(x)} - e^{-u(x)}}{e^{u(x)} + e^{-u(x)}} \quad (1)$$

where

$$u(x) = \frac{(x - \bar{x})}{\sigma_x} \quad (2)$$

is the Z-score that ensures normalization with a mean ( $\bar{x}$ ) of zero and standard deviation ( $\sigma_x$ ) of one.

The expression in Equation (1) can be further simplified by multiplying the right hand side by  $e^x$  and at the same time divides it by  $e^x$  to give;

$$f(u(x)) = \frac{e^{2u(x)} - 1}{e^{2u(x)} + 1} \quad (3)$$

The expression in Equation (3) corresponds to the inverse Fisher transform that has the advantage that it is compressive and for large absolute values, the output is compressed to 1 at most while also removing low amplitude variations. The inverse Fisher transform is analogous to edge sharpening in digital image processing. Moreover, it is the exact solution of the standard Fractional Riccati Differential Equation (FRDE) [45] of the form;

$$D^{(\alpha)} f(t) + y^2(t) - 1 = 0; t > 0, 0 < \alpha \leq 1 \quad (4)$$

where

$$\alpha = 1, f(0) = 0, t = u(x) \quad (5)$$

The values of  $F(u(x))$  lie in the interval  $[-1, 1]$ , but we desire a normalizer that compute values in the interval  $[0, L]$  where  $L$  is the maximum grayscale value such as 255. If we add 1 to both sides of Equation (3), we have:

$$1 + f(u(x)) = \frac{2}{1 + e^{-2x}} \quad (6)$$

To achieve the desire goal of having a normalizer that compute values in the interval  $[0, L]$ , we multiply both sides of Equation (6) by  $L$  and at the same time divide by 2 to have:

$$\frac{L(1 + f(u(x)))}{2} = \frac{L}{1 + e^{-2u(x)}} \quad (7)$$

By substituting Equation (2) into Equation (7) and dividing the left hand side of Equation (7) by  $f(u(x))$ , we have:

$$f(u(x)) = \frac{L}{1 + e^{\frac{-k(x-\bar{x})}{\sigma_x}}} \quad (8)$$

Equation (8) is the desired normalizer, which is a particular form of growth function with  $L$  being the modifier,  $\bar{x}$  is the data mean and  $\sigma_x$  is the data standard deviation while  $k=2,3,4,5 \dots$  and the value of  $f(x)$  lies in the interval  $[0, L]$ .

---

**Algorithm 1** Data pre-processing and transformation for the extraction of features

---

- 1: Read the raw physiological data from the DEAP dataset
- 2: Determine physiological data class using emotion representation
- 3: Preprocess the physiological data by applying inverse Fisher transformation
- 4: Map the transformed physiological data to hyperspectral images
- 5: Extract features from the hyperspectral images using different standard algorithms of digital image processing techniques – HOG, LBP and HIM
- 6: Apply principal component analysis on the extracted features to compute dimensionally reduced Eigen features
- 7: Select the desired Eigen features using the Kaizer criterion of Eigenvalues greater than one
- 8: Use pattern recognizer to recognize the selected Eigen features as low or high - arousal, valence, dominance and liking respectively

This function corresponds to inverse Fisher normalizer with  $k=2$ . After the normalization and applying the inverse Fisher transform on the physiological data which involve mapping the transformed data to grayscale image space, we thereafter form channels' images of each channel's data. This is to enhance the digital image processing techniques which we intend to apply for feature extraction. The subjects' quantitative rating ranges from 1-9 and we thresholded them for each of the valence, arousal, dominance and liking dimensions along two classes of low and high by placing the threshold in the middle as also done in [7] such that all subject's ratings above 4.5 score are respectively mapped to

high (arousal, valence, dominance) and like while ratings below 4.5 score are mapped to the respective low emotion dimensions and dislike representations. Following the procedure shown in Algorithm 1, the Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP) and Histogram of the Images (HIM) features were extracted from the inverse Fisher transformed physiological data. These features are called 'local' features as they were extracted from each channel of each sample. This pre-processing and feature extraction procedure in conjunction with the implementation was done using MATLAB 2018a environment.

### 3.3 Feature Extraction Techniques

Feature extraction is an important task in pattern recognition studies. It involves the use of specific algorithms to acquire discriminating characteristics from raw data for recognition. In this study, we have experimented with the HOG, LBP and HIM feature extraction techniques because they have recorded tremendous success in digital image processing, speech processing, bioinformatics and other pattern recognition applications [23-24, 26-27, 29-31,46-48].

#### 3.3.1 Histogram of Oriented Gradient

The Histogram of Oriented Gradient (HOG) was developed for human recognition and object detection, considering that local object appearance and shape of an image can be represented by the distribution of intensity gradients or edge orientations [46,47]. Since the digitized physiological data have been presented as images, the intensity gradient of the images can be computed to represent discriminatory features. The implementation of the HOG involves dividing an image into cells and compiling the histogram of gradient directions for the pixels within the cells. The aggregation of these histograms represents the HOG features. To compute the HOG feature from a given image, four essential steps are required which are masking, orientation binning, local normalization and block normalization. Research work reported in [46] contains detailed information regarding the computations and characteristics of the HOG features.

#### 3.3.2 Local Binary Pattern

The Local Binary Pattern (LBP) descriptor was originally developed by Ojala, Pietikainen and Harwood [48]. It describes the texture of an image and has been widely applied in diverse applications [48-51]. It assigns numeric label for the block of pixels of an image through a thresholding process that uses a 3x3 neighborhood of the center pixel value while treating the result obtained as a binary number. Since the neighborhoods to the center pixel consist of 8 pixels, the texture descriptor is derived from the histogram of the  $2^8 = 256$  different labels. The LBP value is computed following the steps described in [48]. We chose the LBP descriptor because of its proficiency in appropriately describing the texture of an image [48-52]. In addition, it has a modest theoretical definition which is the foundation of its status as a computationally efficient image texture descriptor [53].

#### 3.3.3 Histogram of Images

Digital image processing involves procedure of obtaining useful information from images by determining an image's pixel properties and variations for the purpose of analysis, classification and recognition or identification. The histogram of an image represent the pixels intensity values in the image. It is a graphical representation that covers all the various intensity values in the image. Thus, after pre-processing and inverse Fisher transformation of the DEAP dataset, the data obtained for each sample is converted and mapped into a grayscale image representation with pixel intensity values ranging from 0-255. As an image processing algorithm, the histogram features representing the pixel intensity values in the various grayscale images is computed using an automatic binning algorithm that yield bins with a uniform breadth which are selected to cover the range of pixel intensity elements thus revealing the underlying unique shape and patterns of the distribution. Histogram as features has strong capabilities for identification and differentiation of patterns and we therefore employ it for human emotion recognition along the valence, arousal, dominance and liking class labels. Histogram based features have been used for image processing and in several pattern classification studies [25, 28] thus necessitating its choice as a third option to the HOG and LBP descriptors employed.

Therefore, the local HOG, LBP and HIM features of each channel for the EEG modality for instance, are reduced from 81, 256 and 256 feature vector sizes respectively to 10, 31 and 31 most dominant features respectively by applying a simple dimensionality reduction algorithm based on the principal component analysis (PCA). The associated eigenvectors are the most dominant or principal component elements of the feature vectors respectively as dimensionally reduced feature vectors are capable of improving classifiers' performance as reported in the literature [54]. Thus, for each sample of the physiological signals to be trained by the RBFANN pattern recognizer, the feature vector sizes of the HOG, LBP and HIM features are now 320, 992 and 992 respectively after concatenating features for all the 32 channels available in each sample of the EEG modality data.

#### 4. Classification Algorithm for Emotion Recognition

Feature classification involves the cataloguing of the extracted features into appropriate classes or states using a pattern recognition or classification algorithm. Such pattern classifier is always trained to learn the inherent characteristics in the different extracted features and attempt to match features with similar patterns into the same class. The RBFNN is the pattern classifier that was utilized for the various experiments conducted in this study.

The RBFANN is a feed forward artificial neural network for solving problems of pattern recognition and function approximation [55,56]. The concepts of RBF are ingrained in earlier pattern recognition techniques such as clustering, spline interpolation, mixture of models and function approximation [55,57]. A typical RBF neural network consists of an input layer, one hidden layer consisting of RBF neurons and an output layer of artificial neuron for each class to be classified [55-58]. Each neuron in the hidden layer implements a radial basis activation function that represent an arbitrary basis for the input vectors, while the network output is a linear combination of radial basis functions of the input and neuron parameters.

According to McCormick [56], classification task performed by a RBFANN measures the input similarity to samples from the training dataset. A “prototype” representing one of the samples in the training dataset is stored in each RBFNN neuron as classification of a new input involves each neuron computing the Euclidean distance between the new input and its prototype. The new input is classified as belonging to Class 1 prototypes if it resembles Class 1 than Class 2 prototypes. The prototypes are indeed cluster centres computed as the average of all the data points in the cluster.

Each of the n-dimensional feature vector which are named, the HOGPEPS, LBPPEPS, HIMPEPS; HOGPS, LBPPS, HIMPS and HOGHES, LBPHEP and HIMHES for the peripherals; EEG and fused modality respectively are extracted from the DEAP dataset and respectively fed into the network through the input layer for classification. Because of its reputation for good performance, the most popular similarity function called the Gaussian radial activation function was employed as a one dimensional input vector in the RBFNN classifier configurations for the experiments conducted in this study. Furthermore, the RBFNN classifier is noted for its ability to approximate continuous functions arbitrarily. It has a faster training process because of its local mapping attribute as compared to other neural networks, very robust to noise [59-61] while it is also capable of yielding at least 10% higher accuracy than can be obtained by the back propagation ANN algorithm [62]. More details about the RBFNN classifier can be obtained in the literature [55,56].

#### 4.1 Experimental Models

The generic architecture for the design and implementation of the procedures for the 3 experimental models consisting of 36 experiments conducted in this study is shown in Fig. 1.

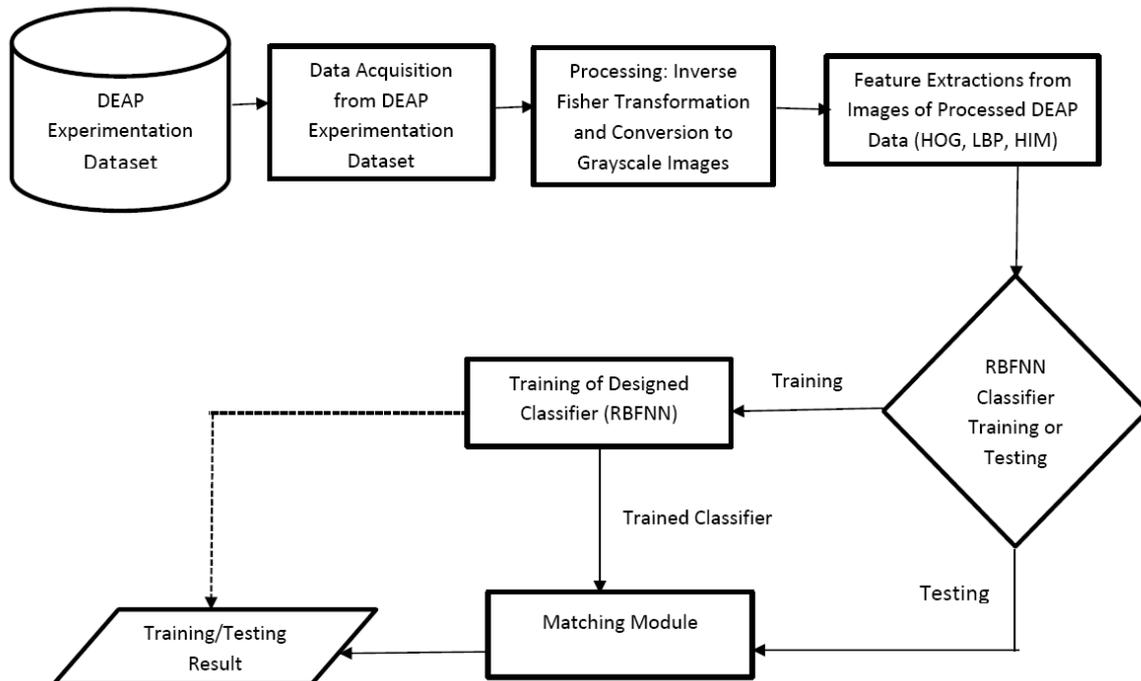


Fig. 1 - The generic architecture of the human emotion recognition experimental model

The first step in the architecture is the DEAP dataset from which the physiological signals as our experimental data was acquired. This step was followed by data pre-processing procedure involving data normalization. Thereafter, the inverse Fisher transform algorithm and mapping of the transformed data to image space was conducted. Feature extractions using the signal and digital image processing algorithms including the HOG, LBP and HIM were subsequently applied and the extracted features were passed to the RBFNN pattern classifier for training or testing and the results of human emotion recognition were obtained.

The first experimental model in this study was conducted on the 8 channels peripheral physiological data consisting of the EOG, EMG, GSR, RESP, BVP and TEMP data of the DEAP dataset from which the HOG, LBP and HIM features named the HOGPEPS, LBPPEPS and HIMPEPS respectively were extracted. There are 1,280 samples in the dataset which was obtained from the 32 subjects' 63s (60 second trial and 3 second pre-trial) duration physiological signals of 40 trials per subject. The twelve experiments performed in this first experimental model was conducted to determine how best the combined peripheral physiological data could accurately recognize human emotions along the arousal, valence, dominance and liking classes. We utilized the HOGPEPS, LBPPEPS and HIMPEPS features for experimentations along each of these four emotion classes or dimensions. The RBFNN classifier was first applied on the arousal class with the HOGPEPS, LBPPEPS and the HIMPEPS features in the MATLAB R2018a environment. For each sample of the 1280 dataset, the HOGPEPS feature vector contain 48 elements per sample of the training data. This serves as the input data to the RBF network and therefore has 48 neurons in the input layer while the output layer has 2 neurons for the 2 emotion classes indicating the low and high arousal binary classes for classification.

In order to train an RBFNN network, determining the number of neurons in the hidden layer is very essential as this affects the result that can be obtained. The prototypes as well as the beta coefficient of the RBF neurons and the matrix of output weights between the RBF neurons and the output node are the parameters that must be carefully selected in the course of determining the number of neurons in the hidden layer [56]. There exist no strict rule in the literature for selecting the prototypes for the RBF neurons. One approach is to create an RBF neuron for each training sample [54] such that for the problem at hand, we would have 1280 neurons; while the other is to randomly select  $k$  prototypes from the training samples [56]. These requirements are slack because with adequate number of neurons, an RBFNN can outline any random complex decision boundary and recognition accuracy can always be improved upon by adding more RBF neurons in the hidden layer. However, a trade-off between the efficiency of the RBF network and the accuracy parameters should be considered because more RBF neurons will indicate more computation cost as it is essential that an excellent accuracy is obtained with the possible minimum number of RBF neurons.

A novel method for selecting the prototypes is to perform k-Means clustering on the training sample while selecting the cluster centres as the prototypes [56]. The average of all the data points in the cluster is computed as the cluster centres. In addition, while utilizing the k-Means algorithm, the training samples are clustered according to classes such that samples from multiple classes are not included in the same cluster.

In order to enhance the network's efficiency and reduce computation costs, instead of using all the available 1280 neurons for the hidden layer, we determined the optimal number of neurons in the hidden layer of our RBFNN by varying the number of clusters between 50-250 per emotion class. For instance, for the arousal class with 2 classes (high or low), with 50 clusters per class, this gives 100 neurons in the hidden layer for the 1280 training samples. We chose 50-250 number of clusters which indicate 100-500 neurons as we leveraged on [54] where the same range of neurons were chosen for each hidden layer in the MLP-ANN configuration consisting of 534 training samples, 2 hidden layer and 14 classes. The authors also utilized all the available 534 neurons in the hidden layer for the RBF network configuration in their study. Our RBF network is therefore much simpler and efficient than the MLP-ANN and RBF configurations reported in [54] as a maximum fewer hidden neurons (39.06%) were utilized out of the available 1280 neurons. The results obtained with the varied number of clusters for the HOGPEPS feature were chronicled. This experiment was further extended to using the LBPPEPS and HIMPEPS features of the peripheral physiological modality and classification was also done along the arousal classes (high or low).

The input feature vectors for the LBPPEPS as well as the HIMPEPS features have 56 elements each, representing the input neurons in the RBF network. The number of neurons in the hidden layer is also experimentally determined as earlier done for the HOGPEPS feature while the number of output neurons remain as 2 representing the arousal classes of high or low.

The first experimental model was concluded by separately utilizing the three features HOGPEPS, LBPPEPS and HIMPEPS extracted from the peripheral physiological modality for classification of human emotion along the outstanding three emotion representation classes namely valence, dominance and liking. The number of input neurons and the features employed are the same with those used for the arousal class but the number of optimal neurons in the hidden layer varies. In all, 12 experiments were conducted in the first experimental model that was mapped to the peripheral physiological modality data.

The second experimental model utilizes the EEG modality data. The inverse Fisher transformed 32 EEG channels of the DEAP dataset was used and the HOG, LBP and HIM feature descriptors was applied to extract corresponding features which was named the Histogram of Oriented Gradient Physiological Signal (HOGPS), Local Binary Pattern Physiological Signal (LBPPS) and the Histogram of Images Physiological Signal (HIMPS) features respectively. The first experiment conducted in this second experimental model is on the arousal class. In classifying the extracted

features into the high or low arousal classes, the HOGPS features have 320 elements as feature vectors for each of the 1280 training samples and was fed into the RBF network. The number of neurons in the input layer is thus 320 while the output layer has 2 neurons. As executed in the first experimental model, the optimal number of neurons in the hidden layer was experimentally determined between 50-250 as we intend to reduce the complexity and computation costs of the RBF network by not utilizing all the available 1280 neurons of the total input samples to obtain the best recognition result.

This experiment was extended to separately using the LBPPS and the HIMPS features of the EEG modality data for the arousal class. The number of neurons in the input layer is 992 each for the LBPPS and the HIMPS features. This huge number of input neurons is as a result of the higher number of channels (32) and the dominant components in the EEG modality data above that of the peripheral physiological data. However, the output layer has 2 neurons consisting of the high or low arousal class. Thus, three sets of experiments yielding three results were performed with the features extracted from EEG modality with the arousal emotion class. In concluding the second experimental model, the outstanding three emotion representation classes of valence, dominance and liking were each and separately used with each of the HOGPS, LBPPS and HIMPS features of the EEG modality data. The same parameters as utilized in the arousal class experiments in the second experimental model are adopted. In summary, 12 experiments were performed in the second experimental model mapped to the EEG modality data.

The third experimental model is the last sets of experiment to be performed. It utilized the fused modality (EEG+peripheral physiological) 40 channels data of the DEAP dataset from which the HOGHES, LBPHEs and HIMHES features were extracted for human emotion recognition task. For the extracted HOGHES features classification into the arousal class, there are 160 elements in each sample feature vectors indicating that the input layer of the RBF network has 160 neurons and the output layer has 2 neurons. The number of clusters or neurons per class was experimentally determined. This experiment was extended to using the LBPHEs and HIMHES extracted features respectively for classifying human emotions along the arousal class. The RBF network has 1440 and 1560 input neurons respectively for the LBPHEs and HIMHES features and the number of optimal neurons in the hidden layer of the RBFNN was experimentally determined.

The valence, dominance and liking emotion representation classes are subsequently used for human emotion recognition with the fused modality data as the HOGHES, LBPHEs and HIMHES features were each and separately used for classification along each of these three emotion classes. In this third experimental model, a total of 12 experiments were performed. The general differences in the three experimental models lies in the modality data used, hence the features extracted, the number of input neurons fed to the network and the optimal number of neurons in the hidden layer with which the best result was attained. The number of output neurons are however the same for all the experimental models as 2 neurons were utilized for each emotion dimension.

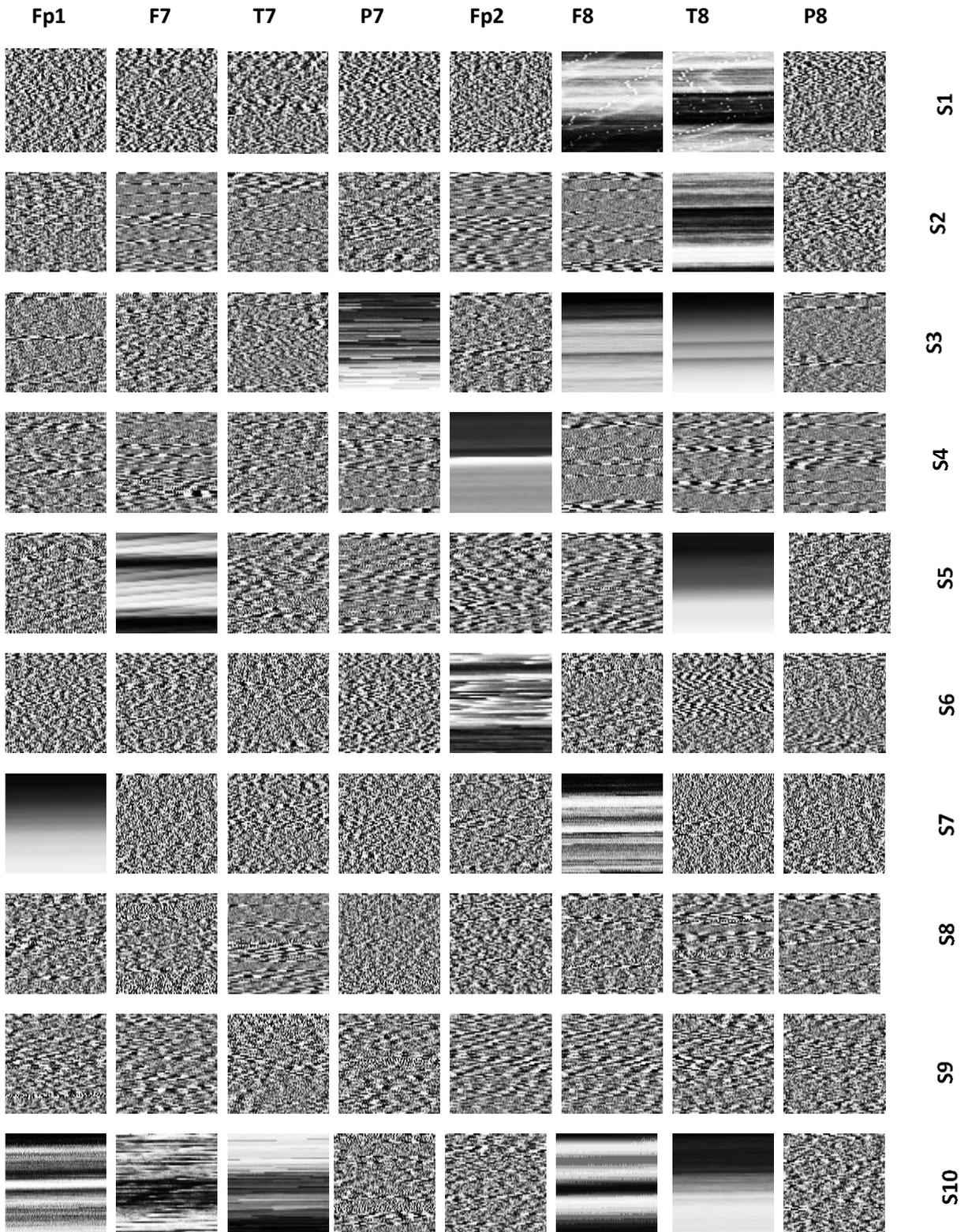
## 5. Experimental Results and Discussion

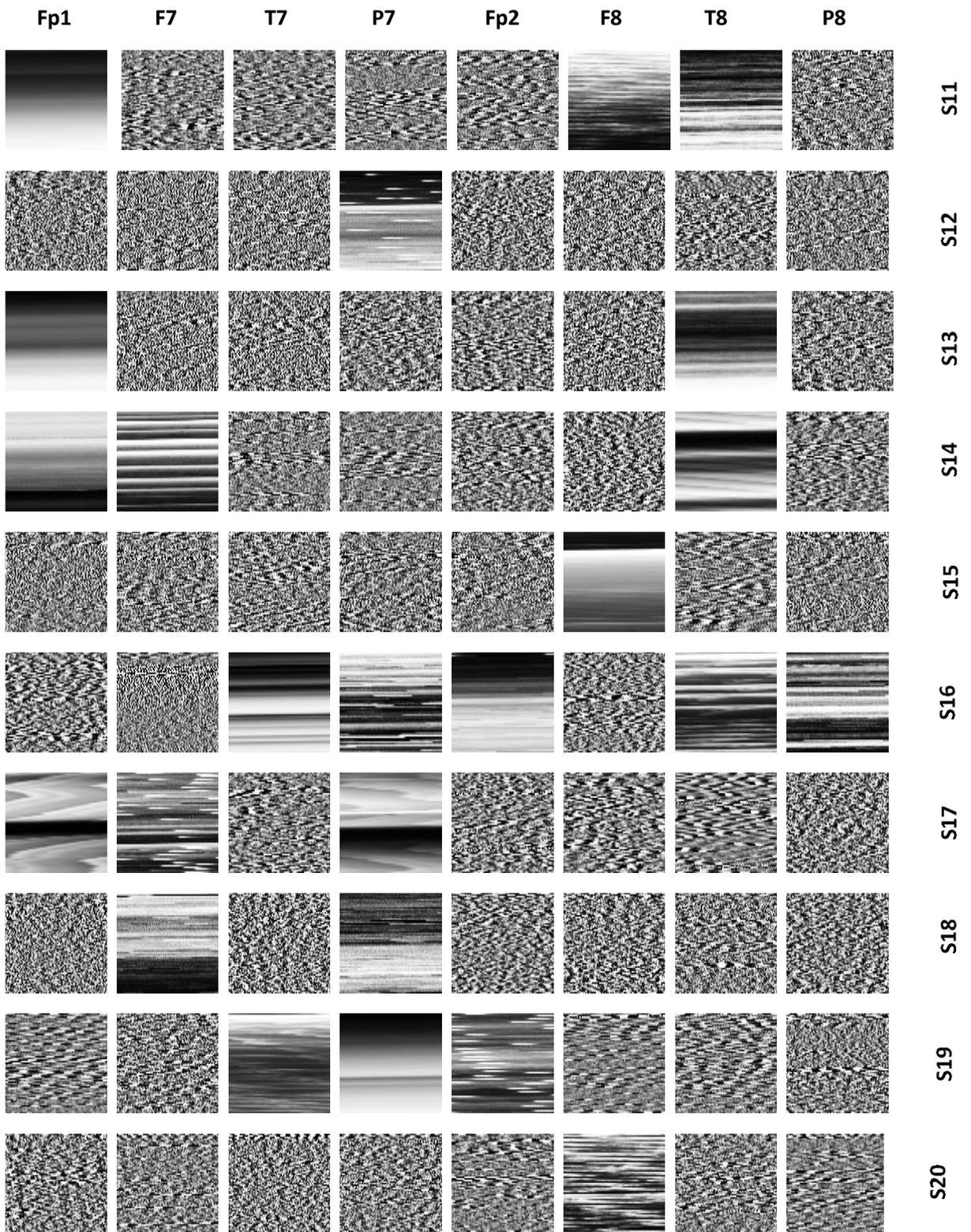
This section presents the results of the 36 experiments carried out in this study. The qualitative results of the inverse Fisher transformation of the emotion physiological data are first presented in order to ascertain the similarities across subjects' emotional responses despite the variation in individual's emotional experiences [63]. Out of the 40 experimental trials of each subject, 8 trials were randomly selected such that the trials selected from each of the 32 subjects (S1-S32) are unique to enable an appropriate trial mix towards enhancing the generalization and reliability of inferences drawn and results obtained. From the 32 EEG channels, the Fp1, F7, T7, P7, Fp2, F8, T8 and P8 channels were identified and selected having been reported in literature to be directly related to human emotions [8, 39, 40, 64-67]. The grayscale image plots of the inverse Fisher transformed physiological data of each subject for 8 random trials and identified channels are shown in Fig. 2 indicating the similarities in the patterns of the images across the subjects. As shown in Fig. 2, there seems to be a similar textural pattern in the majority of the images across the subjects despite the individual variability that do exist in emotional experiences which necessitates the little variations noticed in the few image patterns of one or two channels per subject.

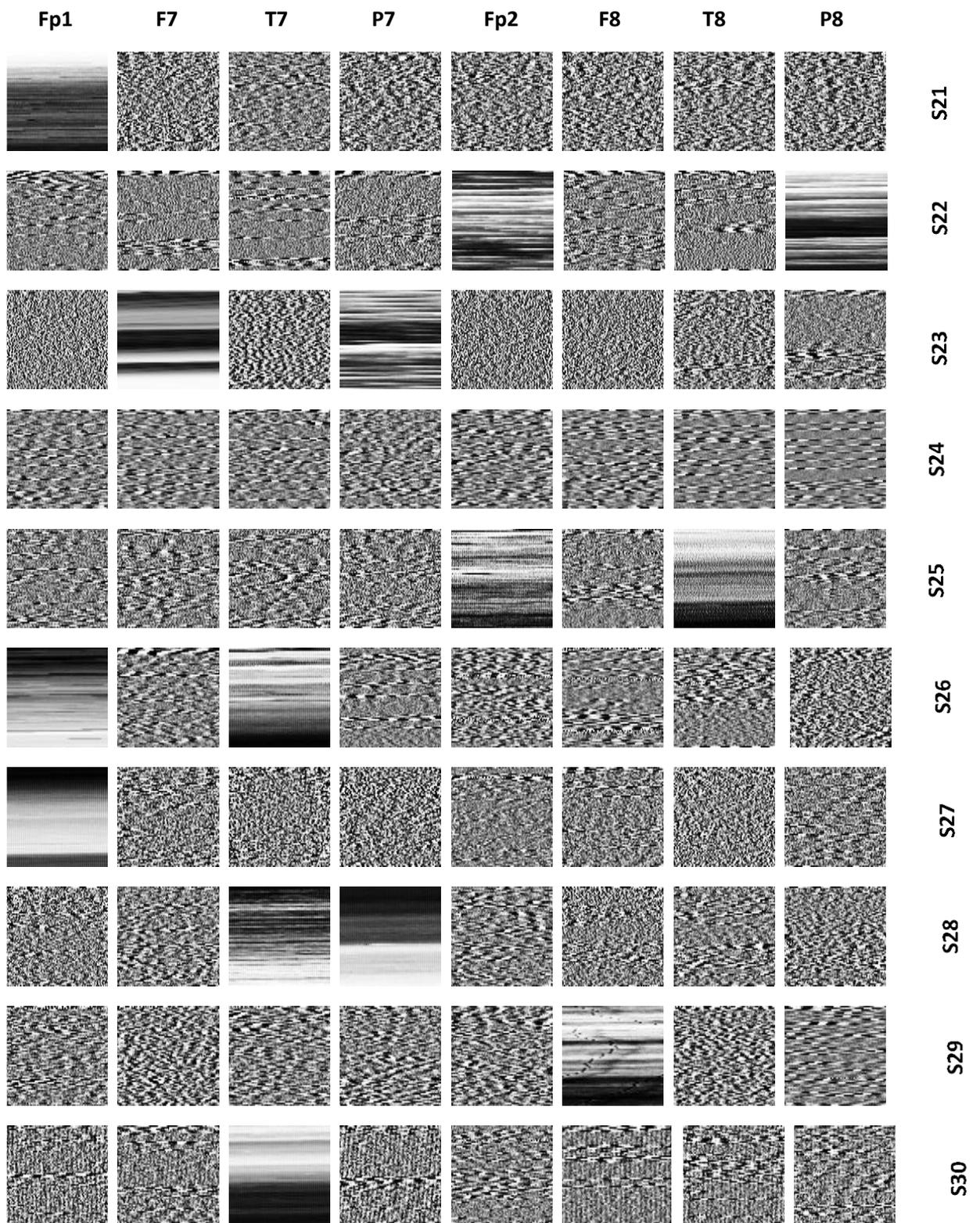
The textural similarity in the images across the subjects is observed in the dense and coarse patterns of the images with observed varied brightness or darkness as a result of the different pixel intensity values and distributions in each image. Thus the strength of the inverse Fisher transformation method that was applied to the physiological data revealed the inherent similarity between the subjects' emotional experiences and responses towards ensuring a subject-independent based inferences and results.

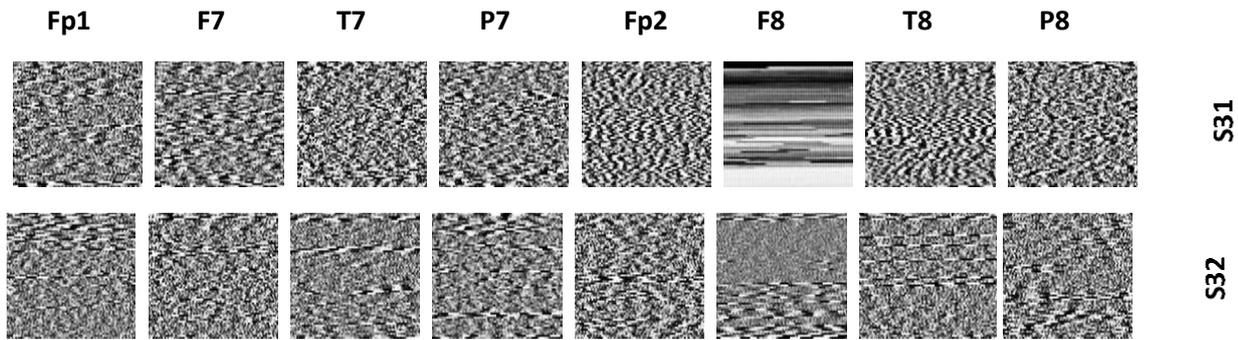
In addition to the grayscale image plot of the transformed physiological data from which the respective features are extracted, the histograms representing accurate distributions of the pixel intensity values for each of the grayscale images in Fig. 2 are investigated. The range of pixel intensity values was binned to 256 that is 0-255, which represent the grayscale image pixel intensity distribution. The importance of the histogram data is that it reveals the density of the underlying distribution of the transformed physiological data and can be employed for probability density function estimation of the underlying variable especially by a pattern classifier. Histograms have been utilized in digital image processing for image analysis, brightness, equalisation, stretching and thresholding. There is no ideal image' histogram shape, but the notable patterns include the unimodal, multimodal, bimodal, skewed right, skewed left and symmetric. In

this study, it was observed that the multimodal pattern dominates the histogram sketches and has multiple peaks which could indicate that the physiological emotional responses of subjects have several patterns of responses and preferences









**Fig. 2 - Images of inverse Fisher transformed EEG physiological data**

that support the literature position that the emotional experiences and responses of different subjects may not be the same [63].

In addition, it was noted that for all the selected channels and across subjects, the pixel count evenly covers a wide range of pixel intensity which indicate a good contrast properties of the image as well as a similarity in the contrast property among the subjects. The shape of the histograms for some of the channels are noticed to be similar for each subject while some similarities can also be observed across many subjects. It is from these pixel intensity values that the respective features are computed for a subject-independent emotion recognition task that is employed in this study.

The extracted features are observed to be similar across subjects as they have the same shape, but usually of varied sizes as a result of the varied feature value otherwise termed amplitude or intensity of the signal. This indicates the robustness of the inverse Fisher transformation method that was applied as well as the feature extraction algorithm in revealing the similarities inherent in a specific feature descriptor of the physiological data of the various subjects.

The aim is to show the differences in shapes and patterns between the different modalities as well as the feature descriptors while also revealing the similarities across the subjects and channels within a particular feature and modality. The strength of the data transformation technique and features employed enhances the attainment of impressive results.

The subjective evaluations carried out are however complemented, as the quantitative analysis and results obtained using the RBFNN pattern classifier and the respective feature descriptors are hereby presented.

For the peripheral physiological modality, the arousal scale results consist of classifying human emotion along the high or low arousal classes. The RBFNN classifier using the HOGPEPS features with 50 clusters per class which represents 100 neurons in the hidden layer, achieved a recognition accuracy of 72.11% with a Mean Square Error (MSE) of 0.7187. The number of clusters which determines the number of neurons that is used in the hidden layer, was varied between 50-250 in a step of 50. It was noticed as shown in Table 1, that as more clusters or neurons are added, the recognition accuracy increases until it peaked at 85.16% with 250 clusters after which the accuracy starts to decline. Thus the best recognition accuracy attained using the HOG features of the peripheral physiological data is 85.16% (MSE=0.0398). This result is indeed very promising and better than the best result of 69.2% obtained with the SVM classifier by [39] that used the bandwaves spectral power density features of the DEAP physiological signals dataset.

**Table 1 - Results of the arousal dimension of the peripheral modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.11	0.7187	73.36	0.3300	73.83	0.2854
100	78.75	0.5874	78.75	0.0630	79.84	0.3828
150	82.27	0.3343	84.22	0.2214	80.23	0.2987
200	84.92	0.1630	84.92	0.4973	82.50	0.2391
250	85.16	0.0398	81.72	0.3667	83.98	0.2089

The LBPPEPS features which represent the LBP features extracted from the peripheral physiological modality data recorded a recognition accuracy of 73.36% (MSE=0.3300) with 50 clusters per class or 100 neurons in the hidden layer. As shown in Table 1, the best recognition accuracy of 84.92% (MSE=0.4973) was however achieved with 200 clusters as the network performance suffers a decline with subsequent increase in the number of hidden layer neurons. However, this result is marginally lower by just 0.24% to the result of 85.16% (MSE=0.0398) posted by the

HOGPEPS. In terms of network’s efficiency, the LBPPEPS result is preferred because, it was achieved with a lower number of neurons in the hidden layer as an increase of 100 neurons yielding just 0.24% marginal accuracy might be considered not efficient.

However, the HIMPEPS features recorded the best recognition accuracy of 83.98% (MSE=0.2089) using 250 clusters or 500 neurons in the hidden layer as shown in Table 1. This result is lower by 1.18% and 0.94% respectively to the results posted by the HOGPEPS and LBPPEPS features. This indicate that for the peripheral physiological signal modality of the DEAP dataset, with our proposed method, the pixel intensity values as features for classification is not better than the HOGPEPS and LBPPEPS features. Despite the seeming lower performance of the HIMPEPS features, the recognition accuracies achieved by the HOGPEPS, LBPPEPS and HIMPEPS features are all better than the arousal class best results of 77.19% [37], 69.2% by the SVM and 74.0% by the random forest classifiers [39], 71.99% [35], and 60.9% [33] recorded by the various research studies.

Furthermore, with the valence class, the HOGPEPS features achieved a recognition accuracy of 85.94% (MSE=0.4246) with 500 neurons in the hidden layer as shown in Table 2. This result is marginally better than the 85.16% (MSE=0.0398) posted by the same features under the arousal class label. In the same vein, a recognition accuracy of 83.20% (MSE=0.3057) was recorded by the LBPPEPS features. This was achieved with 400 neurons in the hidden layer of the RBFNN and the performance declined with more additional neurons. This result is lower than the results of the HOGPEPS features of the valence class, including also the corresponding LBPPEPS result and those of the HOGPEPS and HIMPEPS features of the arousal class.

**Table 2 - Results of the valence dimension of the peripheral modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.58	0.5655	72.66	0.2560	71.88	0.4860
100	79.69	0.3468	79.53	0.4364	77.50	0.0226
150	82.50	0.2009	81.41	0.2704	80.23	0.4129
200	83.83	0.3798	83.20	0.3057	82.66	0.4792
250	85.94	0.4246	82.58	0.3633	82.27	0.4686

As shown in Table 2, a recognition accuracy of 82.66% (MSE=0.4792) was achieved with the HIMPEPS features with the valence class using 400 neurons in the RBFNN hidden layer. The import of the results posted by the various features under the valence class of the peripheral modality data is that, the HOGPEPS features have the most discriminatory properties with which human emotion can be recognised using the DEAP peripheral physiological dataset. This is because, the highest results are posted by the HOGPEPS features across the valence and arousal classes as shown in Tables 2 and 1 respectively, though the 84.92% (MSE=0.4973) recorded by the LBPPEPS features could be preferred, if efficiency is considered above accuracy as a 100 fewer neurons were expended. However, these best results achieved for the peripheral physiological modality using the HOGPEPS, LBPPEPS and HIMPEPS features for the valence class are better than the 69.10% [35], 76.17% [37], 51.2% [33] and 58.2% [34] ] reported in various recent studies that utilized the DEAP dataset with classification done along the valence class.

As contained in Table 3, a classification accuracy of 87.34% (MSE=0.3775) was achieved with 500 neurons in the hidden layer of the RBFNN using the HOGPEPS features with the dominance emotion class. However, with 100 neurons less in the hidden layer, a classification accuracy of 87.03% (MSE=0.1609) was recorded. In terms of comparing efficiency and accuracy, the 87.03% recognition results is preferred because of its associated lesser network complexity and computation cost.

**Table 3 - Results of the dominance dimension of the peripheral modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	73.98	0.4688	75.00	0.0185	74.84	0.1337
100	80.47	0.1096	82.27	0.1766	81.48	0.1377
150	86.48	0.4033	84.14	0.0904	83.52	0.1595
200	87.03	0.1609	84.92	0.1271	83.75	0.1082
250	87.34	0.3775	83.83	0.1528	84.14	0.1700

From the result obtained, the respective HOG features of the high and low dominance classes was further investigated to determine some of the inherent characteristics in the features and classes that could have imparted this impressive result. As shown in Fig. 3, 6 subjects (S4, S8, S12, S18, S22 and S27) were uniformly selected and, the plots of the HOG features for the high and low dominance classes revealed different feature values otherwise termed amplitude or intensity in this study. It was observed that the high dominance class has a higher amplitude than the low dominance class, thereby enhancing easy classification by the pattern recognizer. For instance, amplitude values of 197, 187, 179, 203 and 177 respectively, were recorded for the high dominance class for subjects S8, S12, S18, S22 and S27. Conversely, the low dominance class has amplitudes of 195, 173, 152, 198 and 157 respectively for the subjects and only subject S4 has a higher amplitude for the low dominance class than the high dominance class. However, the trend of a high amplitude value for high dominance dimension and low amplitude value for low dominance dimension is predominant among all the subjects and further contributes to the classification result obtained.

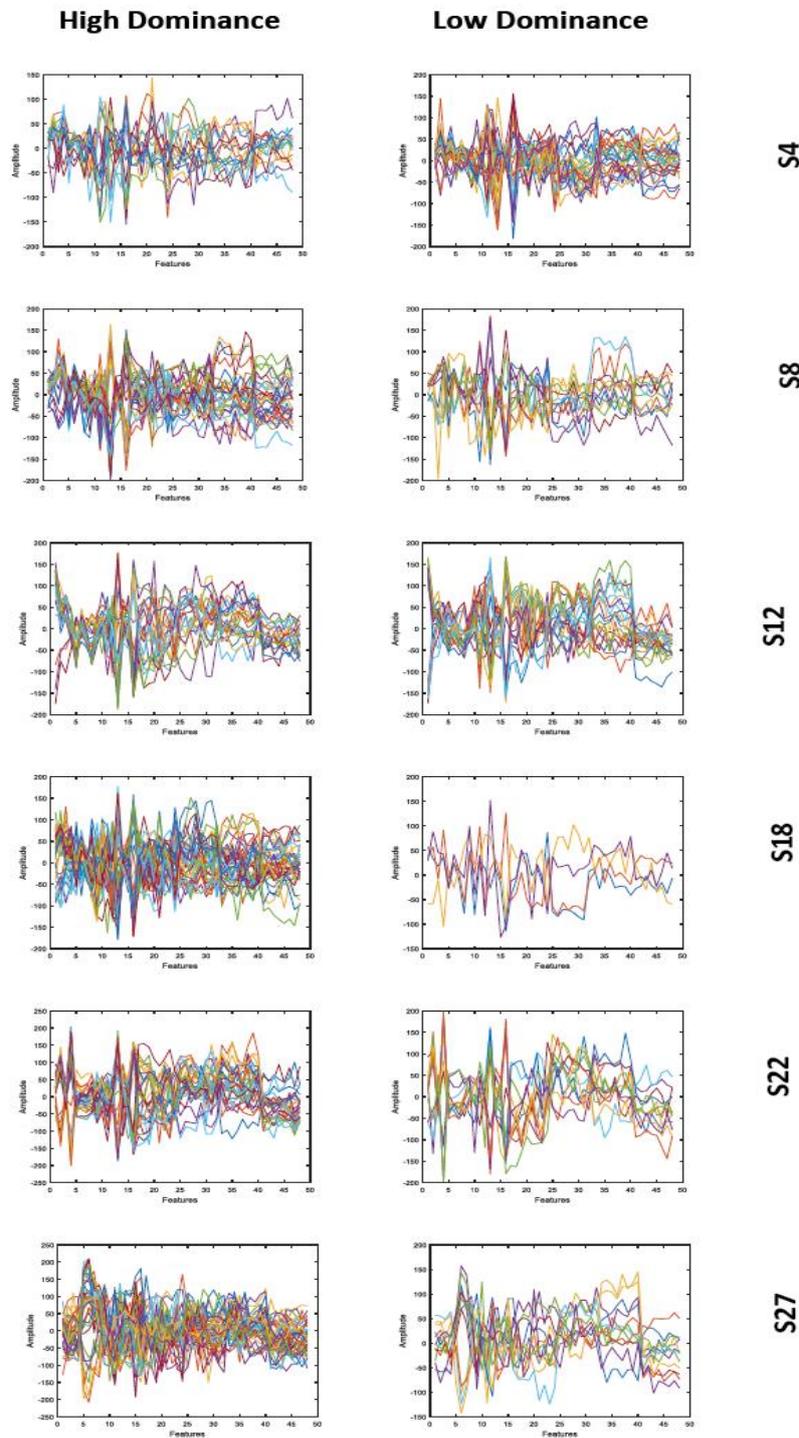


Fig. 3 - Images of sampled HOG features of dominance dimension of peripheral modality data

In emotional signal processing including speech signal, an amplitude is otherwise referred to as intensity and can be computed in various ways which are a measure of the maximum change in a quantity that occurs when the signal is being transmitted [68] as a peak amplitude is a measure of emotional intensity. It has also been established in the literature that as a general rule, the larger the amplitude, the greater is the intensity of the signal [68] including emotional speech and physiological data. In addition, an increase in emotional intensity will trigger an increase in performance up to an optimal point [69] as human emotions are associated with amplitude fluctuations while a highly significant effect of amplitude in various emotions do exist [70-72]. Emotions such as fear, sadness, fear, disgust, joy and boredom have been detected with high, medium and low amplitude values expressed by various subjects [72, 73].

In the next experiment, the LBPPEPS features recorded the best recognition accuracy of 84.92% (MSE=0.1271) using 400 neurons in the hidden layer with the dominance emotion class. However, both the results of 87.34% (MSE=0.3775) and 87.03% (MSE=0.1609) posted by the HOGPEPS features as shown in Table 3 are better than the 84.92% (MSE=0.1271) of the LBPPEPS features and also agrees with the trend earlier recorded in both the arousal and valence classes where the HOGPEPS features also outperforms the LBPPEPS features.

The best recognition accuracy recorded by the Histogram of Images features with the dominance class was 84.14% (MSE=0.1700) while using 500 neurons in the hidden layer of the RBFNN classifier. This performance though falls short of the HOGPEPS and LBPPEPS features' results under the dominance class but it is better than the 83.98% (MSE=0.2089) and 82.66% (MSE=0.4792) respectively recorded by the HIMPEPS features of the arousal and valence classes respectively. So far, the HOGPEPS features and the dominance class is the best combination for the peripheral physiological modality using the DEAP dataset. This clearly indicate that the level of submissiveness of subjects in response to emotional feelings using the 40 emotion elicitation music videos is high and also indicate the ability of the subjects to quantitatively report their emotional feelings. The various best recognition accuracies achieved with the HOGPEPS, LBPPEPS and HIMPEPS features by the peripheral physiological modality with classification done along the dominance dimension are all better than the 65.1% recorded for the dominance dimension by [33] who also used the DEAP dataset.

Liking is the next emotion class that was considered. As shown in Table 4, for the HOGPEPS features, the best recognition accuracy of 83.52% (MSE=0.1647) was achieved with 400 neurons in the hidden layer of the RBFNN as a decline in performance was noticed as more neurons are added. This performance falls short of the results obtained with the same features under the arousal, valence and dominance labels. However, with a cluster size of 150, the LBPPEPS features posted 86.33% (MSE=0.0645) as its best recognition performance which is more than all the results achieved with the respective features under the arousal, valence, dominance and liking class except the 87.34% (MSE=0.3775) result of the HOGPEPS features with the dominance class. This might suggest that more discriminatory and useful information are inherent in this LBPPEPS feature set of the liking dimension for the peripheral modality.

**Table 4 - Results of the liking dimension of the peripheral modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.94	0.5234	75.63	1.0002	74.30	0.6995
100	81.80	0.5475	82.03	0.1173	81.48	0.6488
150	82.58	0.5642	86.33	0.0645	83.20	0.7421
200	83.52	0.1647	82.50	0.3229	83.13	0.5425
250	82.81	0.1808	77.35	0.0155	80.23	0.6789

This LBP result of the liking dimension was further investigated to determine the inherent distinctiveness in the features and class among the subjects that necessitated the recognition's performance. As earlier done with the HOG result of the dominance dimension, the 6 subjects (S4, S8, S12, S18, S22 and S27) uniformly selected posted different feature values otherwise called amplitude for the like and dislike classes. For instance, subjects S4, S8, S12, S18 and S27 posted feature values of 1911, 4278, 4656, 4239 and 694 respectively for the like class. These are higher than the feature values of 1873, 2771, 3368, 4154 and 474 recorded respectively by the subjects for the dislike class, thus revealing the pattern classifier's prowess towards achieving the result obtained. The result aligns with the earlier findings in literature that the larger the amplitude, the greater is the intensity of the signal [68] including emotional speech and physiological data as the like class has a higher intensity than the dislike class.

Furthermore, a recognition accuracy of 83.20% (MSE=0.7421) as shown in Table 4 is the best result posted by the HIMPEPS features and this was achieved with 300 neurons in the RBFNN hidden layer. In addition, all the three best results of 83.52%, 86.33% and 83.20% achieved for the liking dimension of the peripheral physiological modality data

by the HOGPEPS, LBPPEPS and HIMPEPS features respectively are better than the 68.4% and 66.3% classification results achieved by [34,33] along the liking dimension using the DEAP dataset.

The results of the EEG modality experiments are hereby presented. For the arousal class, as shown in Table 5, the HOGPS features yielded a recognition accuracy of 88.28% (MSE = 0.1851) with 500 neurons in the RBFNN hidden layer. This result is better than all the results recorded by all the three different features along all the four emotion dimensions with the peripheral physiological modality data. This confirms the literature position that the EEG modality is capable of producing a better result than the peripheral physiological modality [3]. This performance is also better than the results recently reported in the literature by other authors who also used the DEAP dataset [39,40].

Along the arousal dimension of the EEG modality, the LBPPS features recorded its best recognition result of 83.20% (MSE=0.4066) with the RBFNN using 400 neurons in its hidden layer as shown in Table 5. This performance matches the 83.20% (MSE=0.3057) also reported using the LBPPEPS features for the valence dimension as well as the HIMPEPS features of the liking dimension for the peripheral physiological modality data as shown in Tables 2 and 4 respectively. However, the result falls short of the performance of 88.28% (MSE=0.1851) posted by the HOGPS features indicating that despite its huge feature vector size of 992 elements, the LBPPS features does not contain as many useful and discriminatory data as available in the 320 elements feature vectors of the HOGPS features.

**Table 5 - Results of the arousal dimension of the EEG modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	74.77	0.3817	72.97	0.3212	87.50	0.3590
100	81.56	0.4234	79.22	0.3567	90.70	0.2139
150	85.08	0.2790	81.41	0.4462	92.89	0.1913
200	87.66	0.0975	83.20	0.4066	93.13	0.4188
250	88.28	0.1851	82.73	0.0526	93.36	0.2974

However, as shown in Table 5, the HIMPS features which represent the pixel intensity values of images of the EEG modality data recorded its best recognition accuracy of 93.36% (MSE=0.2974) with 500 neurons in the hidden layer of the RBFNN classifier. This result is very remarkable and surpasses all the results reported for the peripheral physiological modality and that of the HOGPS and LBPPS features of the EEG modality. This is an indication that with our proposed data pre-processing, inverse Fisher transformation and mapping to images technique, the histogram of images is a very potent feature for human emotion recognition as it has also yielded good performances in other studies [25,28]. The 93.36% (MSE=0.2974) recognition result is also better than the results obtained in various recent research studies that have utilized the DEAP dataset [36-37, 39-40].

With this result, the images of Histogram features of the low and high arousal dimensions was analysed with a view to determining some of the inherent characteristics in the features and classes that yielded this result. For the 6 uniformly selected subjects, the plots of the Histogram features for the low and high arousal dimensions revealed different feature values or amplitude. It was observed that across all the selected subjects, the high arousal dimension has a higher amplitude than the low arousal dimension for each of the subjects thereby enhancing easy classification by the pattern recognizer. For instance, amplitudes values of 246, 389, 232, 252, 234 and 460 were recorded for the high arousal class for subjects S4, S8, S12, S18, S22 and S27 respectively while the corresponding amplitude values for the low arousal class are 226, 296, 142, 243, 217 and 313 respectively. This trend of high amplitude for high arousal dimension and low amplitude for low arousal dimension is predominant among the subjects and also agrees with the findings reported for the HOGPEPS and LBPPEPS features of the dominance and liking dimensions of the peripheral modality data.

As shown in Table 6, the RBFNN classifier with the HOGPS features of the EEG modality posted its best recognition result of 88.05% (MSE=0.3022) along the valence dimension while utilizing 500 neurons in its hidden layer. This result is marginally lower than the 88.28% (MSE=0.1851) recorded with the HOGPS features with the arousal dimension but better than the LBPPS features' performance as well as all the results reported under the peripheral physiological modality.

On the other hand, the LBPPS features of the EEG physiological data recorded a recognition accuracy of 85.31% (MSE=0.2523) while 500 neurons was utilized in the hidden layer of the RBFNN. Following a similar trend as noticed in the results earlier reported, the 85.31% (MSE=0.2523) result is lower than the 88.05% (MSE=0.3022) achieved with the HOGPS features using the same number of neurons.

With the HIMPS features, a recognition accuracy of 92.81% (MSE=0.3502) as shown in Table 6 was achieved with 400 neurons in the hidden layer of the RBFNN classifier. An additional 100 neurons was only able to marginally

increase the result by 0.08% to 92.89% (MSE=0.0878). This might not be significant enough, if the efficiency of the network is considered more relevant than the marginal accuracy obtained. Notwithstanding, this highest result obtained with the HIMPS features is better than the result of the HOGPS and LBPPS features of both the arousal and valence dimensions of the EEG modality. In addition, it is marginally lower by 0.47% to the highest result of 93.36% (MSE=0.2974) achieved with the HIMPS feature along the arousal class. Thus, for the EEG modality, the HIMPS features yielded the best recognition result for both the valence and arousal dimensions.

**Table 6 - Results of the valence dimension of the EEG modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	72.58	0.6499	75.31	0.4917	86.64	0.2818
100	81.09	0.4594	81.33	0.2990	91.25	0.1138
150	84.92	0.3973	81.56	0.3877	92.03	0.0685
200	86.48	0.1809	84.84	0.2622	92.81	0.3502
250	88.05	0.3022	85.31	0.2523	92.89	0.0878

The dominance dimension with the HOGPS, LBPPS and HIMPS features posted varied results. The HOGPS features with a feature vector size of 320 and 1280 training samples posted its best recognition accuracy of 89.53% (MSE=0.2345) as shown in Table 7 and was achieved with 500 neurons in the hidden layer of the RBFNN. However, this result is better than those obtained by the HOGPS and LBPPS features for the arousal and valence dimensions of the EEG modality already considered. It is also observed that the results posted by any of the HOGPS, LBPPS and HIMPS features are often better than the corresponding results obtained with the peripheral physiological modality as the ability of the features extracted from the EEG data to outperform the features extracted from the peripheral physiological modality is further proven and aligns with the literature [3].

With the LBPPS features of the EEG modality data classified along the dominance dimension, a recognition accuracy of 87.89% (MSE=0.0430) shown in Table 7 and achieved with 400 neurons of the RBFNN classifier performs less than the corresponding HOGPS features. Though, this result is better than the results of the LBPPS features for the arousal and valence classes, it still confirms that the EEG modality of the DEAP physiological data are more responsive to the HOG than the LBP descriptor. It also shows that the edges and corners of the grayscale images from which these features are extracted are better captured than the local patterns of the images thereby resulting in better HOGPS features' performances than the LBPPS features. While still classifying along the dominance scale, the HIMPS features of the EEG modality achieved its best recognition accuracy of 93.36% (MSE=0.2623) using 300 neurons in its hidden layer as indicated in Table 7. This result is better than the 92.89% (MSE=0.0878) recorded by the same features when classified along the valence dimension while it also matches the 93.36% (MSE=0.2974) obtained with the arousal dimension. The high performance of the HIMPS features of the EEG data for the dominance dimension was also investigated for the 6 subjects uniformly selected for the low and high dominance dimension. This is to determine if there are similar properties inherent in the features and classes among these subjects, which could indicate a subject-independent result as also shown in the arousal dimension of the EEG data modality.

**Table 7 - Results of the dominance dimension of the EEG modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.72	0.6589	76.80	0.8443	86.64	0.6510
100	82.34	0.1717	80.78	0.0454	91.88	0.2910
150	85.86	0.3414	85.47	0.0849	93.36	0.2623
200	88.05	0.2387	87.89	0.0430	93.20	0.3694
250	89.53	0.2345	85.31	0.1099	93.20	0.3068

The amplitude of the high dominance dimension for subject S12 is 119 while the low dominance dimension is 96. Similarly, subjects S18, S22 and S27 have amplitudes of 87, 84 and 107 respectively, for the high dominance dimension with 77, 47 and 88 respectively for the low dominance dimension. With marginal values, only subjects S8 and S4 did not align with this observation. However, the observed trend is predominant across the subjects and the

same as also noticed in the HIMPS features of arousal dimension signals of the EEG modality as the large amplitude values are associated with high dominance dimension and low amplitude values are attached to the low dominance dimension. The few signals that do not follow this trend may be due to some emotions and emotional responses of some subjects which may cover any amplitude range including high, medium and low [73] but these subjects are few out of the 32 subjects considered in the DEAP data set that is utilized in this study.

In another vein, the liking dimension of emotion is utilized with the HOGPS, LBPPS and HIMPS features for classification. According to the various results shown in Table 8, the HOGPS recorded its highest recognition accuracy of 88.83% (MSE=0.1101) with 400 neurons in the hidden layer of the RBFNN classifier. This result is better than all the results posted by the LBPPS for the arousal, valence and dominance dimensions as well as marginally better than the HOGPS results of the arousal and valence dimensions. The best recognition result of 85.23% (MSE=0.1002) was however achieved with the LBPPS features using 400 neurons and with this performance, the HOGPS features are more responsive to and yielded better performance with the RBFNN algorithm than the LBPPS features with the DEAP data set and data transformation method applied.

As the experiment was extended to the HIMPS features and classification done along the liking dimension, a recognition accuracy of 93.13% (MSE=0.1136) was achieved using 300 neurons after which performance declines with further neurons addition. Out of the HIMPS features' results of the EEG modality data for the four emotion dimensions considered, this 93.13% (MSE=0.1136) result is only marginally better than the 92.89% (MSE=0.0878) obtained with the valence dimension. The fact is hereby reiterated that the results obtained along the liking dimensions are not necessarily emotion feelings but subjective ratings of participants indicating their tastes of like

**Table 8 - Results of the liking dimension of the EEG modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.17	0.2599	77.03	0.4403	86.17	0.4480
100	83.28	0.0190	84.22	0.1092	92.19	0.2335
150	86.64	0.0825	84.30	0.3361	93.13	0.1136
200	88.83	0.1101	85.23	0.1002	92.50	0.1253
250	87.89	0.0855	84.84	0.2150	91.95	0.1401

or dislike of the various emotions or elicitation materials. Thus, the liking dimension ratings aside from assisting in tagging of elicitation materials for human emotion recognition can also be used for predictive analytics including addiction management, suicidal thoughts, crime control as well as marketing, advertisement and sales of materials with emotional contents.

The results obtained with the fused modality data using the various features and emotion representation classes are hereby presented. It has been shown in the literature that modality fusion is capable of yielding an improved classification result in a human emotion recognition system [3,7]. This is because complementary characteristics in the different modalities are exploited with a view to detecting a unique pattern from which features are extracted with the primary aim of obtaining a higher performance.

The RBFNN pattern classifier using the HOGHES features for classification along the arousal class achieved a recognition accuracy of 87.11% (MSE=0.2134) with 400 neurons in the hidden layer as shown in Table 9. When compared with the corresponding results of 85.16% (MSE =0.0398) in Table 1 and 88.28% (MSE=0.1851) in Table 5 achieved with the HOG features of the peripheral physiological and EEG modalities data respectively under the arousal dimension, this result is only better than the 85.16% (MSE =0.0398) accuracy of the peripheral physiological modality and falls short of the 88.28% accuracy (MSE=0.1851) of the EEG modality. This result therefore implies that the fused modality did not always lead to a higher performance. This might be traced to the somewhat lower performance of the peripheral physiological data which negatively impact on the 87.11% (MSE=0.2134) result that was obtained. This trend was also shown in the liking dimension F1- score result in [7], where a fusion of EEG, peripheral and Multimedia Content Analysis (MCA) features yielded an F1-score of 61.8% which though more than the 50.2% and 53.8% attained by the EEG and peripheral modalities respectively but falls short of the 63.4% obtained with the MCA features. In addition, the fused peripheral and MCA modalities only yielded an F1-score of 62.2% which is also less than the 63.4% obtained by only the MCA modality. Furthermore, in the same study [7], with an F1-score of 58.3%, 53.3% and 61.8% recorded for the EEG, peripheral and MCA modalities respectively obtained with classification done along arousal dimension, a marginally low 61.6% F1-score was obtained when the three modalities were fused. These results are lower than the best single modalities' results recorded as noted under the arousal and liking dimensions thus agreeing

with our findings as well as literature position that classification results might not always necessarily improve with fused modalities. Notwithstanding this, as shown in [7], the F1-score posted by the fused best two modalities for the arousal and valence dimensions respectively are better than any of the single modality in these dimensions.

With the LBPHEs features of the fused modality data and the arousal dimension scheme, the best recognition accuracy of 81.80% (MSE=0.1968) obtained using 300 neurons in the hidden layer is shown in Table 9. This result is less than the 83.20% (MSE=0.4066) and 84.92% (MSE=0.4973) achieved respectively with the LBPPS and LPBPEPS features as shown in Tables 5 and 1 respectively. Hence the fused modality result achieved does not indicate its superiority to those of the single modalities and it is also more computationally expensive as it contains a huge feature vector size of 1440 elements. The HIMHES result on the hand recorded its best recognition result of 88.36% (MSE=0.3363) with 500 neurons and aligns with the trend earlier noticed in the HOGHES and LBPHEs features wherein the fused modality data does not yield a better result than both results of the single modalities.

**Table 9 - Results of the arousal dimension of the fused modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.47	0.7774	75.00	0.2645	78.52	0.2749
100	82.50	0.2621	76.80	0.2713	82.03	0.4684
150	85.47	0.3005	81.80	0.1968	86.25	0.4769
200	87.11	0.2134	80.16	0.4179	88.13	0.3302
250	87.03	0.0572			88.36	0.3363

The valence dimension is the next scheme that was considered for the fused modality data. The RBFNN classifier was applied on the HOGHES features and 90.08% accuracy (MSE=0.4309) was obtained as the best result as shown in Table 10. The number of neurons utilized in the hidden layer to achieve this best result was 500. This result is better than both the 88.05% (MSE=0.3022) and 85.94% (MSE=0.4246) obtained by the corresponding HOGPS and HOGPEPS features respectively as shown in Tables 6 and 2. This clearly shows that the HOGHES features with classification done along the valence dimension confirms and aligns with the facts in literature that modality fusion can indeed improve performance of a pattern classifier as the feature set is enriched with more useful and discriminatory information [3].

However, with the best recognition result of 80.39% (MSE=0.2964) achieved by the RBFNN using the LBPHEs features with 400 neurons as shown in Table 10, this performance is lower than the 85.31% (MSE=0.2523) and 83.20% (MSE=0.3057) results respectively obtained with the corresponding LBPPS and LPBPEPS features as shown in Tables 6 and 2. This similar trend is also recorded by the HIMHES features where its best recognition accuracy of 88.36% (MSE=0.3970) shown in Table 10 is only able to surpass the 82.66% (MSE=0.4792) of the HIMPEPS features but falls short of the 92.89% (MSE=0.0878) attained by the HIMPS features. Thus, it is only with the HOGHES features of the valence dimension that the modality fusion approach improved the classification result obtained by the RBFNN pattern classifier.

**Table 10 - Results of the valence dimension of the fused modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	75.47	0.6702	75.16	0.7735	80.31	0.4552
100	80.78	0.3733	78.91	0.7226	84.06	0.4287
150	86.72	0.3506	79.77	0.3857	85.31	0.4795
200	88.28	0.3864	80.39	0.2964	88.28	0.4678
250	90.08	0.4309	79.84	0.3807	88.36	0.3970

As observed in Table 11, the dominance emotion dimension, using the HOGHES features of the fused modality data yielded the best recognition accuracy of 88.36% (MSE=0.1396) using 400 neurons in the hidden layer of the RBFNN. The 89.53% accuracy (MSE=0.2345) in Table 7 and 87.34% accuracy (MSE=0.3775) in Table 3 as best results obtained with the HOGPS and HOGPEPS features respectively with 500 neurons each indicates that the fused modality features performance cannot be a replacement for the EEG single modality.

**Table 11 - Results of the dominance dimension of the fused modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.09	1.1032	76.95	0.2945	82.19	0.0179
100	79.69	1.1365	81.33	0.2092	86.17	0.0648
150	85.39	0.1590	80.86	0.3424	89.77	0.1183
200	88.36	0.1396	80.78	0.3559	87.97	0.1108
250	86.88	0.0843	78.36	0.4133		

This trend is also demonstrated with the LBPHEs features' performance wherein the best recognition accuracy of 81.33% (MSE=0.2092) shown in Table 11 achieved with 200 neurons is below the 87.89% (MSE=0.0430) and 84.92% (MSE=0.1271) respectively achieved with the corresponding LBPPS and LBPPEPS features. In the same vein, the HIMHES features with 89.77% best recognition accuracy and MSE of 0.1183 as shown in Table 11, was only able to surpass the 84.14% recognition accuracy and MSE=0.1700 as shown in Table 3 of the HIMPEPS features but falls short of the 93.36% (MSE=0.2623) of the HIMPS features as shown in Table 7; thus cannot replace this single modality performance.

The fused modality data, the HOGHES, LBPHEs and HIMHES features and liking dimension is the next scheme and the various results are shown in Table 12. The HOGHES features have its best recognition result of 88.59% (MSE=0.3303) which was attained using 400 neurons in the hidden layer of the RBFANN. When compared with the performances of the corresponding HOGPS and HOGPEPS features of the single modalities that recorded 88.83% (MSE=0.1101) and 83.52% (MSE=0.1647) respectively as shown in Tables 8 and 4, only the peripheral physiological modality was surpassed by the fused modality result while it falls marginally below the EEG modality result by 0.24% hence not a replacement for the single modalities.

**Table 12 - Results of the liking dimension of the fused modality data**

Features Number of Centres	HOG		LBP		Histogram	
	ACC (%)	MSE	ACC (%)	MSE	ACC (%)	MSE
50	76.80	0.2948	79.77	0.3613	82.58	0.8574
100	83.36	0.3549	79.60	0.0486	84.61	0.2087
150	86.72	0.3729	81.17	0.0252	87.03	0.2613
200	88.59	0.3303	81.41	0.0683	89.22	0.0795
250	86.64	0.2224	80.31	1.0126	89.38	0.0826

Also, with the LBPHEs features, a similar trend is noticed where the best recognition result of 81.41% (MSE=0.0683) shown in Table 12, of the fused modality data cannot match the 85.23% (MSE=0.1002) and 86.33% accuracies (MSE=0.0645) of the EEG and peripheral physiological single modalities shown in Tables 8 and 4. On the other hand, the HIMHES features with its best result of 89.38% (MSE=0.0826) is below the highest performance of 93.13% (MSE=0.1136) of the HIMPS as shown in Table 8 and, only better than the 83.20% of the HIMPEPS features as shown in Table 4, which aligns with the earlier position that the fused modality scheme cannot always replace the single modalities scheme in terms of classification performance.

The summary of the best results obtained under each modality and the corresponding features while using our proposed methods are shown in Table 13. The arousal class recorded 85.16% (MSE=0.0398) as its best recognition accuracy for the peripheral physiological modality using the HOGPEPS features while the EEG and the fused modalities recorded their best recognition results of 93.36% (MSE=0.2974) and 88.36% (MSE=0.3363) respectively using the HIMPS and HIMHES features respectively. Thus based on the results obtained, the peripheral physiological data appears to be easily detected by local appearance and shape of their mapped images as provided by the HOG feature descriptors. However, for the arousal class and across all the modalities considered, the HIM features and EEG modality produced the best result of 93.36% (MSE=0.2974) and this combination is hereby recommended.

As shown in Table 13, for the valence class and across the three modalities, the best recognition results of 92.89% (MSE=0.0878) was achieved by the EEG modality using the HIM features. The fused modality's result of 90.08% (MSE=0.4309) and 85.94% (MSE=0.4246) of the peripheral modality were achieved with the HOG features thus confirming their unique strength of shape and local appearances. Because of its superior performance, a combination of the EEG modality and the HIM features is recommended for the valence class in human emotion recognition using the DEAP dataset. However, the dominance class recorded recognition accuracies of 89.77%

**Table 13 - Summary of experimental results across dimensions, modalities and features**

Emotional Dimension	Modalities					
	Peripheral		EEG		Fused	
	ACC (%)	Feature	ACC (%)	Feature	ACC (%)	Feature
Arousal	85.16	HOG	93.36	HIM	88.36	HIM
Valence	85.94	HOG	92.89	HIM	90.08	HOG
Dominance	87.34	HOG	93.36	HIM	89.77	HIM
Liking	86.33	LBP	93.13	HIM	89.38	HIM

(MSE=0.1183) and 87.34% (MSE=0.3775) as its best results, for the fused and peripheral modalities respectively with the HIM and HOG features. Following the trend noticed in the arousal and valence classes, the best result of 93.36% (MSE=0.2623) was recorded by the EEG modality using the HIM features and the combination is thus recommended. In summary, as contained in Table 13, even for the liking class, the best result of 93.13% (MSE=0.8827) was recorded by the EEG modality using the HIM features. The overall best recognition result is 93.36% which was obtained by arousal and dominance classes but the dominance class has a lower MSE of 0.2623 as the dominance class, EEG modality and HIMPS features combination is hereby recommended for emotion recognition using the DEAP dataset.

To enable replicability of the RBF neural networks experimentations in this study, the parameters that were optimized to obtain the recognition results include the numbers of neurons that was utilized in the hidden layer which was experimentally determined for the HOG, LBP and HIM features respectively for the various modalities. There are 2 neurons in the output layer, which denote the number of emotional states to be classified. The hidden layer and output layer implemented a hyperbolic tangent sigmoid transfer function which is considered simple and suitable above the rectified linear unit (ReLU) of multi-layer deep neural.

## 6. Conclusion

In this study, we have successfully carried out comparative experiments on the suitability of using the HOG, LBP and HIM features extracted from the DEAP physiological data with the RBF-ANN classifier to recognize human emotional states along the arousal, valence, dominance and liking classes. The unique method of our data pre-processing and the application of the inverse Fisher transform algorithm on the preprocessed physiological data greatly enhances the discriminatory strength of the extracted features of the HOG, LBP and HIM descriptors.

According to the literature, emotional physiological data exhibit significant non-linearity [74] as most physiological data do not increase linearly even for only a single subject while experiencing an emotional state. We leveraged on the peculiar non-linearity of emotional physiological data and the inherent capability of neural network models to perform efficiently when the data distribution is non-linear to achieve higher emotion recognition accuracies in the study at hand as the models were able to approximate arbitrary non-linear functions [75]. This is in addition to the model's strength of tacitly detecting and extracting intricate non-linear relationships between independent and dependent variables as well as their capabilities of detecting every probable interaction between predictor variables [76].

To navigate the computational complexity drawbacks in deep learning and explore the possibilities of getting promising results, the RBFANN was applied to the extracted HOG, LBP and HIM features in the experiments conducted. From the summary of results shown in Table13, it can be observed that the HOG, LBP and HIM features offered competitive and varied performances. This indicates that the three feature extraction techniques performed differently on the DEAP dataset by extracting different discriminatory features from it. The various best results obtained by each of the three modalities and features are state of the art and better than the results obtained in recent research studies [7, 33-40, 77-89] that utilized the DEAP dataset. These results are also better than those reported by [34, 36-38] despite the trending deep learning approaches applied in those studies.

Thus, the three set objectives of this study have been successfully achieved. These objectives include, (i) to acquire physiological data from the DEAP corpus for the purpose of human emotion states recognition (ii) to extract discriminatory features from the physiological data using the Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP) and Histogram of Images (HIM) feature extraction techniques and (iii) to compare the performances of the RBFANN models using the three different features for the task of human emotion recognition. In conclusion, based on our findings and results obtained in this study, we recommend a combination of HIM features and EEG modality data with classification done along the dominance scale in order to carry out emotion recognition task using physiological signals. This combination gave the best result of 93.36% (MSE = 0.0153) when compared to the other features sets, modalities and configurations. The other results obtained in this study are also better than the results reported in similar recent research works that utilized the same dataset [33-40]. With the current result, there is a huge prospect of the adoption of an automated emotion recognition paradigm in application domains such as customer services, call centers, safe driving, biomedical and healthcare engineering, emergency services and psychiatric disorder

research. Future works will include the use of other machine learning methods, merging of audio and physiological signals, the discovery and adoption of other physiological feature extraction algorithms and the practical prototyping for real-time application purpose.

## Acknowledgement

The authors would like to thank the Research and Postgraduate Support Directorate, Durban University of Technology, Durban for supporting this research study.

## References

- [1] Maaoui, C., & Pruski, A. (2010) Emotion recognition through physiological signals for human-machine communication, *Cutting Edge Robotics*, Vedran Kordic (Ed.)
- [2] Zhihong, Z., Pantic, M., Glenn, I.R & Thomas, S.H. (2009). A Survey of affect recognition methods: audio, visual and spontaneous expressions, *IEEE Transactions on pattern analysis and machine intelligence* 31(1), 39-58
- [3] Soleymani, M., Jeroen, L., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging, *IEEE Transactions on affective computing* 3(1), 42-55
- [4] Heng, Y.P., Lili, N.A., Alfian A.H., & Puteri, S.S. (2013). A study of physiological signals-based emotion recognition systems, *International journal of computer and technology* 11(1), 2189-2196
- [5] Chun-yan, N., Hai-xin, S., & Wang, J. (2013). The relationship between chaotic characteristics of physiological signals and emotion based on approximate entropy, *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, Atlantic Press, 552-555
- [6] Chanel, G., Kronegg, J., Grandjean D., & Pun, T. (2006). Emotion assessment: arousal evaluation using EEG and peripheral physiological signals. *Multimedia content representation, classification security* (4105), 530-537
- [7] Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok, L., Yazdani, A. , Ebrahimi, T. , Pun, T., Nijholt, A. & Patras, I. (2012). DEAP: A database for emotion analysis using physiological signals, *Affective Computing* 3(1), 18-31
- [8] Noppadon, J., Setha, P. & Pasin, I. (2013). Real-time EEG based happiness detection system, *The scientific world journal*, Hindawi Publishing Corporation, (article id 618649)
- [9] Wagner, J., Jonghwa K. & Andre, E. (2005). From physiological signals to emotion: Implementing and comparing selected methods for feature extraction and classification, *Proceedings of the IEEE international conference on multimedia & expo (ICME)*
- [10] Eun-Hye, J., Byoung-Jun, P., Sang-Hyeob, K. & Jin-Hun, S. (2012). Emotion recognition by machine learning algorithms using psychophysiological signals, *International Journal of Engineering and Industries* 3(1), 55-66.
- [11] Ekman, P. (1982). *Emotion in the human face*, Second ed., Cambridge Univ. Press
- [12] Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacyoyanni-Tarlatzis, I., Heider, K. & Tzavaras, A. (1987). Universals and cultural differences in the judgment of facial expressions of emotion, *Journal of personality and social psychology* 53(4), 712-717
- [13] Hanjalic, A., & Xu, L.Q. (2005). Affective video content representation and modeling, *IEEE Transactions on multimedia* 7(1), 143-154
- [14] Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011). Physiological signals based human emotion recognition: A review, *Signal Processing and its Applications (CSPA)*, *Proceedings of IEEE 7th International Colloquium*, 410-415
- [15] Jing, C., Bin, H., Yue, W., Philip, M., Yongqiang, D., Lei, F., & Zhijie, D. (2017). Subject-independent emotion recognition based on physiological signals: a three stage decision method, *BMC Medical Informatics and Decisions Making* 17(3), 167
- [16] Xiang, L., Dawei, S., Peng, Z., Yazhou, Z., Yuexian, H. & Bin, H. (2018). Exploring EEG features in cross-subject emotion recognition, *Frontiers in Neuroscience* 12(162)
- [17] Yuan-Pin, L., & Tzyy-Ping, J. (2017). Improving EEG-based emotion classification using conditional transfer learning, *Frontiers in Human Neuroscience* 11(334)
- [18] Rigas, G., Katsis, C.D., Ganiatsas, G. & Fotiadis, D.I. (2007). A user independent, biosignal based emotion recognition method, *User Modeling, Lecture notes in Computer Science* (4511), 314-318
- [19] Chien, B.C., Hong, T.P., Chen, S.M., Ali, M., Mohammad, Y. & Nishida, T. (2009). Measuring naturalness during close encounters using physiological signal processing, *Next generation applied intelligence* (5579), 281-290
- [20] Cong, Z., & Chetouani, M. (2009). Hilbert-Huang transform based physiological signals for emotion recognition, *International symposium on Signal processing and Information Technology (ISSPIT)*, 334-339
- [21] Panagiotis, C.P., & Leontios, J.H. (2010). Emotion recognition from EEG using higher order crossings, *IEEE Transactions on Information Technology in Biomedicine* 14(2), 186-197

- [22] Vyzas, E., & Picard, R.W. (1999). Offline and online recognition of emotion expression from physiological data. In proceedings of workshop on emotion-based agent architectures, third international conference on autonomous agents, 135-142
- [23] Hamed, M., Sazzad, H.M., & Rafael, A.C. (2014). Using remote heart rate measurement for affect detection, Proceedings of the twenty-seventh international Florida artificial intelligence research society conference, 118-123
- [24] Happy, S.L., & Aurobinda, R. (2015). Automatic facial expression recognition using features of salient Facial patches, IEEE Transactions on Affective Computing 6(1), 1-12
- [25] Iman, B., Norihide, M., Hiroshi, M., Hasan, D. & Gholamreza, A. (2017). Histogram-based feature extraction from individual gray matter similarity-matrix for Alzheimer's disease classification, Journal of Alzheimer's disease 55(4), 1571-1582
- [26] Manar, M.F.D., Aliaa, A.A.Y. & Atallah, H. (2014). Spontaneous facial expression recognition based on histogram of oriented gradients descriptor, Computer and Information Science 7(3), 31-37
- [27] Ramchand, H., Narendra, C. & Sanjay, T. (2013). Recognition of facial expressions using Local Binary Patterns of important facial parts, International Journal of Image Processing (IJIP) 7(2), 163-170
- [28] Thamizhvani, T.R., Hemalatha, R.J., Babu, B., Dhivya, A.J.A., Joseph, J.E., & Chandrasekaran, R. (2018). Identification of skin tumours using statistical and histogram-based features, Journal of Clinical and Diagnostic Research 12(9), 11-15
- [29] Uroš, M., & Božidar, P. (2015). Automated facial expression recognition based on histograms of oriented gradient feature vector differences, Journal of Signal Image and Video Processing (SIViP), 9(1), 245-253
- [30] Wang, X., Jin, C., Liu, W., Hu, M., Xu, L., & Ren, F. (2013). Feature fusion of HOG and WLD for facial expression recognition, In IEEE/SICE International Symposium on System Integration (SII), 227–232
- [31] Yunan, Z., Yali, L. & Shengjin, W. (2015). Facial expression recognition using coarse-to-fine classifiers, Computer Science and Applications–Hu (Ed), 127-13
- [32] Jun-Wen, T., Adriano, O.A., Hang, L., Steffen, W., David, H., Stefanie, R., Kerstin, L., Holger, H. & Harald, C.T. (2016). Recognition of intensive valence and arousal affective states via facial electromyographic activity in young and senior adults, PLoS ONE 11(1)
- [33] Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks, International Journal of Information and Education Technology 3(5) 2013, 505–511
- [34] Li, X., Zhang, P., Song, D., Yu, G., Hou, Y. & Hu, B. (2015). EEG based emotion identification using unsupervised deep feature learning, Proceedings of SIGIR2015 Workshop on Neuro-Physiological Methods in IR Research, Santiago, Chile. NeuroIR'15
- [35] Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., & Yan, B. (2017). Emotion recognition from EEG signals using multidimensional information in EMD domain, BioMed Research International, (article ID 8317357)
- [36] Li, Y., Huang, J., Zhou, H., & Zhong, N.(2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks, Applied Sciences, 7(1060), 1-20
- [37] Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning, Computer Methods and programs in Biomedicine, 140 (2017), 93–110
- [38] Alhagry, S., Fahmy, A.A., & El-Khoribi, R.A.(2017). Emotion recognition based on EEG using LSTM recurrent neural network, International Journal of Advanced Computer Science and Applications 8(10), 355-358
- [39] Menezes, M.L.R., Samara, A., Galway, L., Sant'Anna, A., Verikas, A., Alonso-Fernandez, F., Wang,
- [40] H. & Bond, R. (2017). Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset, Personal and ubiquitous computing 21(6), 1003–1013
- [41] Nakisa, B., Rastgoo, N.M., Tjondronegoro, D. & Chandra, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors, Expert Systems with Applications 93(2018), 143-155
- [42] Chubb, H., & Simpson, J.M. (2012). The use of Z-scores in paediatric cardiology, Journal of Annals of Pediatric Cardiology, 5(2), 179–184
- [43] Colan, S.D. (2013). The Why and How of Z Scores, Journal of the American Society of Echocardiography 26(1), 38–40
- [43] Singh, Y.N., & Gupta, P. (2007). Quantitative evaluation of normalization techniques of matching scores in multimodal biometric systems, S.-W. Lee and S.Z. Li (Eds.): ICB Lecture Notes in Computer Science, (4642), 574–583
- [44] Raschka, S. (2014). About feature scaling and normalization – and the effect of standardization for machine learning algorithms, [https://sebastianraschka.com/Articles/2014\\_about\\_feature\\_scaling.html](https://sebastianraschka.com/Articles/2014_about_feature_scaling.html)
- [45] Salehi, Y., & Darvishi, M.T. (2016). An investigation of fractional Riccati differential equations. Optik – International Journal for Light and Electron Optics 127(23)

- [46] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection, *Proceedings of the IEEE Computer Society, International Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, United States, (1), 886-893
- [47] Dipankar, D. (2014). Activity recognition using histogram of oriented gradient pattern history, *International journal of computer science, engineering and information technology* 4(4), 23-31
- [48] Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions, *Pattern Recognition* 29(1), 51–59
- [49] Ojala, T., & Pietikäinen, M. (1999). Unsupervised texture segmentation using feature distributions, *Pattern Recognition* 32(3), 477-486
- [50] Ojala, T., Valkealahti, K., Oja, E., & Pietikäinen, M. (2001). Texture discrimination with multidimensional distributions of signed gray-level differences, *Pattern Recognition* 34(3), 727–739
- [51] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multi resolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions Pattern Analysis and Machine Intelligence* 24(7), 971-987
- [52] Garcia-Olalla, O., Alegre, E., Fernandez-Robles, L., Garcia-Ordas, M.T., & Garcia-Ordas, D. (2013). Adaptive local binary pattern with oriented standard deviation (ALBPS) for texture classification, *EURASIP journal on image and video processing* 1(31), 1-11
- [53] Rahim, M.A., Hossain, M.N., Wahid, T., & Azam, M.S. (2013). Face recognition using Local Binary Patterns (LBP), *Global Journal of Computer Science and Technology Graphics & Vision* 13(4), 1-9
- [54] Adetiba, E., & Olugbara, O.O. (2015). Improved classification of lung cancer using radial basis function neural network with affine transforms of voss representation, *PLoS ONE* 10(12)
- [55] Bors, A.G. (2001). Introduction of the radial basis functions (RBF) networks, *The York research database. Online symposium of Electronics Engineers*
- [56] McCormick, C. (2013). Computer vision and machine learning projects and tutorials, *Radial Basis Function Network (RBFN) Tutorial*
- [57] Ugur, H. (2004). Artificial neural networks, Ch.9, *EE543 Lecture notes*
- [58] Xianhai, G. (2011). Study of emotion recognition based on electrocardiogram and rbf neural network, *Procedia Engineering* 15(2011), 2408–2412
- [59] Andina, D., & Pham, D.C. (2007). Computational intelligence for Engineering and Mathematics, Andina D and Pham DC Eds. Springer Science & Business Media, 120-121
- [60] Cha, I., & Kassam, S.A. (1995). Channel equalization using adaptive complex radial basis function network, *IEEE Journal on Selected Areas in Communications* 13(1), 122-31
- [61] Leonard, J.A., & Kramer, M.A. (1991). Radial basis function networks for classifying process faults, *IEEE Control Systems Magazine* 11(3), 31-8
- [62] Chapman, R.A., Norman, D.M., Zahirniak, D.R., Rogers, S.K., & Oxley, M.E. (1991). Classification of correlation signatures of spread spectrum signals using neural networks, *Proceeding of IEEE National Aerospace and Electronics Conference* (1), 485-491
- [63] Siemer, M., Mauss I., & Gross, J.J. (2007). Same situation—different emotions: how appraisals shape our emotions, *Emotion* 7(3), 592–600
- [64] Davidson, R.J., Jackson, D.C., & Kalin, N.H. (2000). Emotion, plasticity, context and regulation: Perspectives from affective neuroscience, *Psychological Bulletin* (126), 890-909
- [65] Niemic, C.P. (2002). Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion, *Journal of Undergraduate Research* (1), 15-18
- [66] Noppadon, J., Setha, P., & Pasin, I. (2013). Emotion classification using minimal EEG channels and frequency bands, *Proceedings of the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE '13)*, , 21–24
- [67] Petrantonakis, P.C., & Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings, *IEEE Transactions on Information Technology in Biomedicine* 14(2), 186–197
- [68] Glenn, E. (2010). The physics hypertextbook: intensity vs amplitude, *Physics info*
- [69] Nasoz, F., Alvarez, K., Lisetti, C. & Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies, *Cognition, Technology & Work* 6(1), 4–14
- [70] Burkhardt, F. (2005). Emofilt: the simulation of emotional speech by prosody-transformation, *Proceedings of Interspeech '05, Lisbon, Portugal*, 509-512
- [71] Hammerschmidt, K., & Jurgens, U. (2007). Acoustical correlates of affective prosody, *Journal of Voice* 21(5), 531-540
- [72] Hartmut, R.P., & Christian, K. (2008). Amplitude and amplitude variation of emotional speech, *Proceedings of annual conference of the Interspeech '08, Brisbane, Australia*, 1036-1039
- [73] Lech, M., Stolar, M., Bolia, R., & Skinner, M. (2018). Amplitude-frequency analysis of emotional speech using transfer learning and classification of spectrogram images, *Advances in science, technology and engineering systems journal* 3(4), 363-371

- [74] Healey, J. (2014). Physiological sensing of emotion, In Calvo R, D’Mello S, Gratch J, Kappas A. (Eds.), *The oxford handbook of affective computing*, Oxford University Press
- [75] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems* (2), 304–314
- [76] Jack, V.T. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, *Journal of Clinical Epidemiology*, 49(11), 1225-1231
- [77] Arnau- González, P., Arevalillo- Herráez, M., & Ramzan, N. (2017). Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals, *Neurocomputing*, 244(2017), 81- 89
- [78] Bahari, F., & Janghorbani, A. (2013). EEG- based emotion recognition using recurrence plot analysis and K nearest neighbor classifier. In *proceedings of 20th Iranian Conference on Biomedical Engineering (ICBME 2013)*, Tehran, Iran.
- [79] Chen, J., Hu, B., Wang, Y., Moore, P. , Dai, Y., Feng, L., & Ding, Z. (2017). Subject-independent emotion recognition based on physiological signals: a three-stage decision method. *BMC Medical Informatics and Decision Making*, 17(3), 167.
- [80] Chen, M., Han, J., Guo, L., Wang, J., & Patras, I. (2015). Identifying valence and arousal levels via connectivity between EEG channels. In *proceedings 2015 International Conference on Affective Computing and Intelligent Interaction (ACII 2015)*, Xi’an, China.
- [81] Choi, E.J., & Kim, D.K. (2018). Arousal and valence classification model based on Long Short-Term Memory and DEAP Data for mental healthcare management. *Healthcare Informatics Research*, 24(4), 309–316
- [82] Chung, S.Y., & Yoon, H.J. (2012). Affective classification using Bayesian classifier and supervised learning, *Proceedings of 12th International Conference on Control, Automation and Systems (ICCAS 2012)*, JeJu Island, South Korea
- [83] Liu, J., Meng, H., Li, M., Zhang, F., Qin, R. & Nandi, A.K. (2018). Emotion detection from EEG recordings based on supervised and Unsupervised dimension reduction, *Concurrency and Computation Practice Experience*, (30)
- [84] Liu, W., Zheng, W.L., & Lu, B.L. (2016). Emotion recognition using multimodal deep learning. In *Proceedings: International Conference on Neural Information Processing (ICONIP 2016)*, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, D. Liu (eds). *Lecture Notes in Computer Science*, Springer, Cham, (9948)
- [85] Naser, D.S., & Saha, G. (2013). Recognition of emotions induced by music videos using DT- CWPT. *Proceedings of 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT 2013)*, Kharagpur, India. 2013
- [86] Qiu, J., Li, X., & Hu, K. (2018). Correlated attention networks for multimodal emotion recognition. In *proceedings IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2018)*, Madrid, Spain
- [87] Rozgic, V., Vitaladevuni, S.N. & Prasad, R. (2013). Robust EEG emotion classification using segment level decision fusion. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1286–1290
- [88] Tang, H., Liu, W., Zheng, W., & Lu, B. (2017). Multimodal emotion recognition using deep neural networks. In *proceedings International Conference on Neural Information Processing (ICONIP 2017)*, D. Liu et al. (Eds.), Part IV, LNCS 10637, 811–819
- [89] Zhuang, X., Rozgic, V. & Crystal, M. (2014). Compact unsupervised EEG response representation for emotion recognition. In *proceedings of IEEE- EMBS International Conference on Biomedical and Health Informatics (BHI 2014)*, Valencia, Spain