Sādhanā (2020)45:166
https://doi.org/10.1007/s12046-020-01354-w

Moth Monarch Optimization-Based Deep Belief Network in Deception Detection System

NIDHI SRIVASTAVA* and SIPI DUBEY

Department of Computer Science Engineering, Rungta College of Engineering and Technology, Bhilai (C.G.), India
e-mail: srivastavanidhi223@gmail.com; drsipi.dubey@gmail.com

MS received 27 September 2019; revised 4 February 2020; accepted 16 March 2020

Abstract. Deception is the action of causing a person to believe something, which is known to be lying with the provision of evidence to support such false beliefs with certain intensions. Identification of the deceptive characteristics manually is a challenging problem for the researchers. Thus, an automatic deception detector is necessary to be developed in order to ensure higher accuracy. Accordingly, this paper proposes a novel deception detector method called Moth Monarch optimization-based Deep Belief Neural Network (MMO-DBN). The proposed MMO-DBN classifier undergoes the phases of feature extraction and classification. Initially, the input speech signals are pre-processed to remove the noise present in the signal and subjected to feature extraction to extract the significant features, such as Mel Frequency Cepstral Coefficients (MFCC), Spectral Kurtosis, Spectral Spread, Spectral Centroid, minimum blood pressure, maximum blood pressure, respiration rate, and Tonal Power Ratio. Then, these extracted features are subjected to classification using Deep Belief Neural Network (DBN), which is trained with the proposed Moth Monarch optimization (MMO) algorithm that is the integration of Monarch Butterfly Optimization (MBO) and Moth Search (MS) algorithm. The performance of the proposed MMO-DBN is analyzed using the metrics, namely accuracy, sensitivity, and specificity. The proposed method obtained the higher accuracy, sensitivity, and specificity of 0.984, 0.9836, and 0.9375, respectively that shows the superiority of the proposed MMO-DBN in deception detection.

Keywords. Detection; deep learning; speech signal; deception optimization; frequency-based features; Moth search; Monarch Butterfly optimization; Mel frequency cepstral coefficients; spectral kurtosis; spectral spread; spectral centroid; tonal power ratio.

1. Introduction

Lie detection, normally called deception detection, makes use of the questioning methods to find truth and falsehood in the response of people. Deception is termed as the active communication of messages and data to produce a false conclusion [1]. Due to large amounts of textual data that are transmitted using Computer-mediated Communication (CMC), people stay highly inefficient and unsuccessful in finding the messages, which are deceptive [2]. In addition, with an increasing number of multimodal communications, the requirement arises for enhanced methodologies for the detection of deceptive characteristics [3]. There are number of methods available on which polygraph test is the commonly used method that can differentiate the truth and the lie. In general, the question–answer method is used with the technique that stores the physiological values in order to differentiate the truth and the lie. During the session of interview, behaviour of the subject is needed to be monitored closely to identify the fluctuations in each response for the identification of the lie syntax. The physiological responses including blood pressure, respiratory rate, and the heartbeat, can be used for differentiating the truths and lie [4–6]. The speech signals are used to detect the deception [7], as the speech utterances provide the data related to emotions [8]. The discounted {0-1} Knapsack problem (DKP) is solved by using 10 effective binary MS-based algorithms, such as MS1-MS9, and MS [9]. MBO algorithm is used for addressing various tasks of global optimization and possessing improved fitness towards accelerating convergence. MBO algorithm based on opposition-based learning (OBL), and random local perturbation (RLP) which can reduce a sorting operation and enhance the computational efficiency [10].

In addition to the advances in deception detection, research related to automatic deception detection with the speech signals has obtained considerable attraction in the recent years. Studies related to deception detection have performed from a psychological perspective with the consideration of multiple indicators, namely biometric, visual, and physical cues [11, 12]. For speech-based deception

*For correspondence
Published online: 27 June 2020

detection, the acoustic-prosodic features are utilized to find the differences among the lie and the truthful speech due to the fact that the energy, pitch, speaking rate and other stylistic factors differs with the speakers deceive [13]. The majority of data are exchanged on daily basis involving some deceit level and is performed with the rich media by using the deep learning approach [14, 15]. Thus, the research has mainly aimed on richly mediated communication channels and the speech signals. It is necessary to know the advantages and the limitations of detecting the deception over less rich mediums, such as e-mail systems. Previous attempts on the detection of deceit were normally performed with the use of the physiological sensors and the trained experts. However, a major limitation of these techniques is that the judgment of human on various cases is normally biased and produced less accuracy in classification. In addition, these methods need more time and effort to perform the analysis [3].

In order to analyze the performance of models involved in detecting deception, various intelligent methods were developed for lie detection. The advanced machine learning algorithms with a number of modalities, such as speech acts as one of the advanced techniques of deception detection. In [16], the requirement of the machine intelligence was analyzed for lie detection to recognize the human affective state. Pattern classification method was used for the differentiation of deception and non-deception using optimization techniques [17, 18]. In [19], cross-cultural deception detection was performed with the collected features using various approaches, such as context-free grammars (CFG), part-of-speech (PoS) tags, and unigrams, for lie detection. The physiological bio-signals, including Galvanic Skin Response (GSR), Electrocardiogram (ECG), and functional MRI (fMRI) are used for the detection of lies immediately. The physiological responses are processed with the use of neural networks, data mining, and statistical analysis methods to obtain the statement veracity [20]. Linear Discriminant Analysis (LDA) [21] is used to find the lie using the P300 frequency and skin pattern (EOG). In [22], a fuzzy theory was used, and the Support vector networks are commonly used in different areas, which is useful in removing the repetitive and irrelevant features [23]. In [24], a new multimodal deception dataset with Decision Trees (DT) and Random Forest (RF) based on visual and textual cues were used. In [25], a manual annotation was used, which was effective in the optimisation of deception detection technique.

The main intention of this research is to design a method for detecting the lie using deep learning by proposing an optimization technique. The proposed method comprises two phases, such as feature extraction phase and the classification phase. In the first step, the input speech signals are pre-processed to remove the noise present in the signal. Then, the pre-processed signal is fed to feature extraction to extract the necessary features, namely MFCC, Spectral Kurtosis, Spectral Spread, Spectral Centroid, minimum blood pressure, maximum blood pressure, respiration rate, and Tonal Power Ratio. These highly significant features are given to the classification module for deception detection with the use of DBN that uses the newly designed MMO optimization algorithm, which is the integration of MBO [26, 27] and MS algorithm [28]. MMO is similar to the MS; the only difference is that the position update equation of MS is modified with the position update equation of MBO. In MMO, the positions of the moth individuals are updated based on the position update rules of MS and the MBO.

The main contribution of the paper is:

Application of the MMO-DBN classifier for the deception detection using the features of the speech signal: The features of the speech signal are responsible for effective classification and enhanced accuracy. Thus, the frequency-based signal features, such as MFCC, Spectral Centroid, Spectral Kurtosis, Spectral Spread, and Tonal Power Ratio. In addition, the additional features, such as minimum blood pressure, maximum blood pressure, respiration rate, and heartbeat rate are considered that play a vital role in the classification process. Finally, the MMO-DBN classifier is used for the recognition of deception of speech among various speakers. The MMO classifier is the integration of the MBO and MS algorithms for deriving the weights of the DBN network for deception detection.

This paper is organized as follows: Section 1 details the need for deception detection. Section 2 provides the literature review. Section 3 deals with the MMO-DBN classifier for deception detection. Results and discussions are discussed in section 4, and finally section 5 explains the conclusion of the paper.

2. Motivation

In this section, the literature reviews of number of methods used for deception detection with their limitations are discussed.

2.1 *Literature review*

In this section, the review of eight existing works related to deception detection is discussed: Srivastava and Dubey [1] developed the ANN and SVM-based method for better pointer deception detection. The limitation of this method was the provision of slower results, the inability to analyze other speech, and physical values. Gogate *et al* [29] developed a deep learning driven multimodal fusion approach for better accuracy, but it was not suitable for advanced deep learning approaches. Abouelenien *et al* [3] designed the multimodal approach that was potentially good for deceptive behaviour detection and performance gain, but this method was not able to consider larger datasets in the presence of variety of new topics, scenarios, and failed to make use of the additional modalities, such as visual features. Zhou *et al* [30] developed the Relevance Vector Machine (RVM) classification method for very fast calculation with smaller memory. This method obtained higher accuracy and robustness with the property of generalization, but it required advanced functions for the enrichment of the detection model. Abouelenien *et al* [31] developed the multimodal deception detection approach, which was a better indicator for the detection of determining the deception with improved accuracy, but it was not suitable to consider the different data for better rate of deception detection. Amanda Chow and John Louie [11] developed the recurrent neural network (RNN), a machine learning method that potentially converged to a higher level of accuracy, but cannot consider the lexical features. O'Shea *et al* [32] modelled the automatic deception detection system that provided dynamic responses to the inputs of the user and had the ability to simulate the complex signals, but lacked the datasets for diverse population representation and the optimization of the neural network classifiers. Nasri *et al* [33] modelled the ReGIM-Lab Lie Detection System, which was very efficient in the classification of the speech signals, but required multimodal system for the provision of the relationship among stress, voice, heart rate and brain activity. James OrShea *et al* [34] developed an automated deception detection system for facial objects detection, and extraction of non-verbal behaviour in the form of microgestures over short periods of time. The problem in this method was, more data capture in diverse population representation, and neural network classifiers optimization. Douglas P Twitchell *et al* [35] developed the automated deception detection systems for contextual information such as base rate and cost, but it had lack of information security.

2.2 *Challenges*

The various challenges of this research are detailed below:

- In [1], the physiological and speech features were utilized for the detection of deception. The speech features in addition to the physical features act as the better pointer for the detection of deception. The main limitation of this method is that the judgement of human provides poor accuracy in classification and needs more time and effort for the analysis.
- The conventional methods of lie detection include the observation on facial expression and emotion. However, these methods have their own limitation due to the fact that it can be faked by the respondent. In addition, the approaches on lie detection basically depend on specialized environmental or instrumentation conditions that may require more time and cost to provide questionable results [20].
- The major limitation noticed in the data driven research on deception detection is the unavailability of real data and true motivation, while analyzing the deceptive behaviour. Due to artificial setting, the subjects may not arouse emotionally that leads to difficulty to generalize the findings to real life scenarios [24].
- The level of textual communication obtained and stored by the individuals and organizations is increasing significantly, due to the expansion of the Internet. Because of the massive rate of textual data transmitted using CMC, people remain unsuccessful in the detection of those messages, which are deceptive [2].
- The challenge faced by the automatic deception detection methods is the production of corpuses. The traditional corpuses rely on artificially collected data, where the subjects are requested to develop stories in both truthful and deceptive ways that lacked the true emotions and the real-world evidence [29].

3. Proposed method of deception detection using deep belief network

The increasing number of multimodal communications has lead to the need for enhanced methodologies to detect the deceptive characteristics. This paper aims to develop a method of deception detection using deep learning with an optimization technique. The proposed method undergoes two phases, such as feature extraction phase and classification phase. Initially, the input speech signals are preprocessed and fed to feature extraction for the extraction of required features. The features, such as frequency-based speech features and the individual-based features are used for deception detection. The frequency-based speech features include the MFCC, Spectral Centroid, Spectral Kurtosis, Spectral Spread, and Tonal Power Ratio, whereas the individual-based features include the minimum blood pressure, maximum blood pressure, respiration rate, and heartbeat rate. These highly significant features are fed to the classification module for the deception detection for which the DBN is used. The DBN uses the newly designed optimization algorithm, named as MMO algorithm, which is the integration of MBO and MS algorithm. Thus, the proposed MMO-DBN performs the deception detection with the use of the obtained features. Figure 1 models the block diagram of deception detection system using the proposed MMO-DBN.

3.1 *Pre-processing of speech signal*

In general, the signal is very sensitive to various environmental factors, and in the same way, the speech signal is affected with various noises. The noise effects that are present in the signal are lowered with the pre-processing process that acts as the initial step in the process of

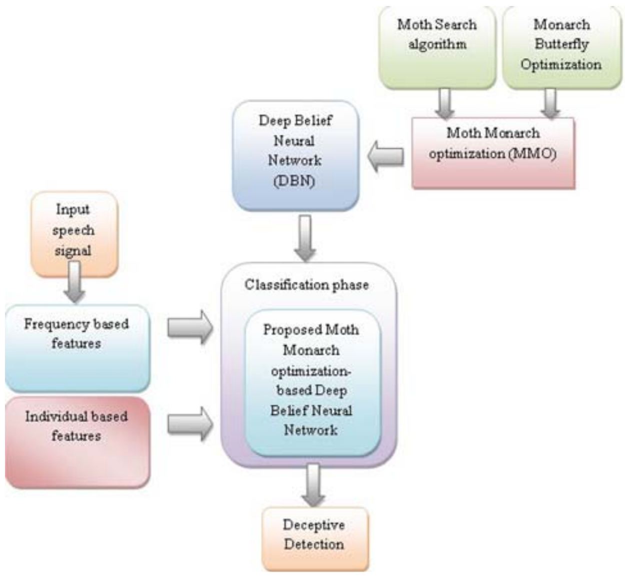

Figure 1. Block diagram of proposed MMO-DBN deception detection system.

deception detection. The relation of the speech signal is expressed as

$$B = \{D(s)\,;\quad 1 \le s \le H\} \tag{1}$$

where, B is the speech database that comprises a set of speech signal in the presence of noise, $D(s)$indicates the speech signals, and H represents the total data samples present in the input signal. After pre-processing the input signal, the enhanced speech signal is obtained and is indicated as $E(s)$.

3.2 *Extraction of features*

The features such as frequency-based features from the speech signal of number of individuals and the individual-based features are subjected to classification process for the detection of deception.

3.2.a *Frequency-based signal features:* The frequency-based features are obtained from the speech signals of individuals, who are suspected under deception. The pre-processed speech signal E(s) is fed to the feature extraction process for the extraction of features, such as MFCC, Spectral Centroid, Spectral Spread, Spectral Kurtosis, and Tonal Power Ratio to perform the classification process of the deception detection system.

a) *Mel Frequency Cepstral Coefficients (MFCC)*

The MFCC [36] acts as the mostly accepted feature representation technique for large number of speech recognition systems. The cepstral coefficients of lower order are very sensitive to the overall spectral slope, whereas the cepstral coefficients of higher order are susceptible to the noise signal. MFCC is a quantitative representation of speech signal and can be mathematically expressed as,

$$MFCC_s = DCT(\log(|FFT(E(s))|)) \tag{2}$$

where, DCT is the discrete cosine transform, and FFT is the Fast Fourier transform. The size of the MFCC is obtained as $[1 \times 14]$.

b) *Spectral Centroid*

The spectral centroid [37] is one of the most commonly used timbre parameters. The centroid is normalised using the highest rate-map centre frequency for the reduction of the effect of gammatone parameters. The relation of the spectral centroid is expressed as

$$S_c = \frac{\sum_{r=1}^{R} U_u[r] * r}{U_u[r]} \tag{3}$$

where, $U_u[r]$ represents the magnitude of the Fourier transform at frame u and frequency bin r. The centroid is the rate of spectral shape and the higher value of centroid is responsible for brighter textures with increased frequencies. The spectral centroid obtained is of the size $[1 \times 1]$.

c) *Spectral Kurtosis*

The spectral kurtosis [38] in the frequency domain is termed as the kurtosis of the frequency components and is normally used for the detection of speech signals that occur in random. The kurtosis $p(u)$ at each frequency bin u is expressed as the spectral kurtosis of $K(e)$ and is expressed as

$$C_u(e) = \frac{C_4\{K^+(e), K^+(e), K^+(e), K^+(e)\}}{[C_2\{K^+(e), K^+(e)\}]^2} \tag{4}$$

where, $K^+(e) \in \{K(e), K^+(e)\}$, $K(e)$ represents the N-point discrete Fourier transform (DFT), $K^+(e)$ represents the complex conjugate of $K(e)$, C_2 is the second order cumulant, C_4 is the fourth order cumulant. The size of the spectral kurtosis is given as $[1 \times 1]$.

d) *Spectral Spread*

The spectral spread [39] represents the average deviation of rate-map over its centroid that is commonly related to the signal bandwidth. As same as the centroid, the spectral spread is usually normalised with the highest rate-map centre frequency in such a way that the feature value lies between zero and one. The spectral spread is otherwise termed as the spectral standard-deviation, and it is expressed as,

$$SS(u_l) = \left(\sum_{v=1}^{V} (w_r - S_c(u_l))^2 \bullet N_r(u_l)\right)^2 \tag{5}$$

where, $SS(u_l)$ is the spectral spread at the frame u_l, $S_c(u_l)$ is the spectral centroid at the frame u_l, v is the sinusoidal harmonic partial, $N_r(u_l)$ is the normalized value of the magnitude STFT, and w_r is the frequency of the bin r. The size of the spectral spread is obtained as [1 × 1].

e) *Tonal Power Ratio*

The tonal power ratio is defined as the ratio of sum of the peak values, which is greater than the constant threshold to the sum of the signal inputs. The tonal power ratio is expressed as,

$$T_{PR} = \frac{\sum_{s=1}^{K} K(s)}{\sum_{s=1}^{E} E(s)} \qquad \textit{if}\ K(s) > \text{Constant threshold} \tag{6}$$

where, $K(s)$ is the sum of peak values that are greater than that of the constant threshold, and $E(s)$ is the sum of the input signals. The obtained size of the tonal power ratio is [1 × 1].

3.2.b *Individual-based features:* In addition to the frequency-based features, the individual-based features such as minimum blood pressure, maximum blood pressure, respiration rate, and heartbeat rate are used for classification in such a way that the classification accuracy is enhanced.

a) *Minimum and maximum Blood Pressure*

The change of temperature [40] in facial areas is a result of change in blood pressure rate. With the assumption of metabolic heat factor being negligible, the thermodynamic equation related to blood pressure is obtained as,

$$\frac{dB_P}{dt} = \frac{\omega_B\left(\left(\mu_c + \varepsilon/3d_p\right) - \mu\right)}{\left(\omega_B - \omega_S\right)} \frac{d\omega_s}{dt} \tag{7}$$

where, B_P is the blood pressure rate, ω_B is the blood temperature at core level and is equal to 310°K, ω_S is the skin temperature, ε represents the thermal conductivity and is equal to 0.168 kcal, d_p is the core temperature depth from the surface of skin, and μ is the constant. When the flow of blood is faster, the blood pressure is maximal, and when the flow of blood is slower, the blood pressure is minimal. The size of minimum and maximum blood pressure feature is obtained as [1 × 1].

b) *Heartbeat rate and respiration rate*

The respiration rate and the heartbeat rate can be successfully extracted from the signals that are received from distances of up to two meters from the subject or person under test. These rates can be used as an effective feature in the detection of deception. The size of the feature obtained for heartbeat rate and the respiratory rate is obtained as [1 × 1].

Feature vector from speech signal: The features combine together to form the feature vector, and it is expressed as,

$$W = \{W_1, W_2, W_3, W_4, W_5, W_6, W_7, W_8, W_9\} \tag{8}$$

where, W_1 refers to the MFCC feature, W_2 is the Spectral centroid feature, W_3 is the Spectral kurtosis feature, W_4 is the Spectral spread feature, W_5 is the tonal power ratio feature, W_6 is the minimum blood pressure feature, W_7 is the maximum blood pressure feature, W_8 is the respiration rate feature, and W_9 is the heartbeat rate feature. The feature vector of size [1 × 22] is generated and this feature vector is representing the frequency and individual-based signal features, which is then subjected to the classification process for the detection of deception.

3.3 *Deception detection using the MMO-based deep belief neural networks*

The extracted feature from the speech signal undergoes the classification using the DBN classifier. The role of DBN [41] is to extract and recognize the patterns present in input speech signal. DBNs are trained to classify the data with the outputs obtained from the past records to produce enhanced accuracy in classification. The classification is performed to identify the deception using the speech signal of the persons, for which the features are obtained from the input speech signal and the individuals and are analyzed using the MMO-DBN classifier.

3.3.a *Proposed Moth-Monarch Optimization algorithm:* The proposed MMO-DBN is used to train the DBN to detect the deception using the features of speech signal obtained from speakers, and the individual-based features. In order to perform this action, the concept of MBO is integrated within the MS algorithm, and thus Eq. (15) obtained is used to train the DBN. The advantage of using the MMO-DBN classifier is that it combines the advantages of both the MS and the MBO algorithms while training the DBN classifier. The biggest advantage of using the MS is that it searches the best solutions more effectively with increased accuracy, whereas the advantage of using MBO is that it suits well for parallel processing and has the capability of developing the trade-off among diversification and intensification. The benefit of DBN classifier includes the effective generation of solutions for the optimization problems. In addition, the DBN classifier can be used in various engineering applications to provide numerous benefits to the industries. Hence, the MMO-DBN classifier is used in the detection of deception with speech signal as input.

Algorithmic steps of the proposed MMO-DBN algorithm: The steps involved in the proposed MMO-DBN algorithm are detailed as:

i) *Initialization:* The initialization of individuals involved in optimal solution generation is the first step of this algorithm. The generation number is set as $x = 1$, and the population of the moths is initialized in random. The index γ, acceleration factor η, maximum generation $J_{\max}$, which is the maximum generation considered as the termination criterion, and the maximum walkstep $L_{\max}$ are initialized.

ii) *Fitness evaluation:* The fitness is fully based on the minimal value of error that is estimated between the classifier output and the estimated output. The moths those are closer to the light sources in the population are the moths with higher fitness and they fly around the best moth in Levy flight form. This characteristic is described in Eq. (9) based on the position of the moth as,

$$P_{i,j}^{x+1} = P_{i,j}^{x} + \beta L_F(\kappa) \tag{9}$$

where, $L_F(\kappa)$ is the step drawn from Levy distribution, and β is the scale factor, the value of which relies on the optimization problem, and is expressed as,

$$\beta = \frac{\chi_{\max}}{x^2} \tag{10}$$

where, $\chi_{\max}$ is the maximum walk step.

iii) *Position update using MMO-DBN algorithm:* The position of the moth individuals is updated based on the position update rules of MS and the MBO. The standard equation of the MBO is expressed as,

$$P_{i,j}^{x+1} = P_{i,j}^{x} + \beta\left(dP_j - 0.5\right) \tag{11}$$

This equation can be rewritten as,

$$P_{i,j}^{x} = P_{i,j}^{x+1} - \beta\left(dP_j - 0.5\right) \tag{12}$$

where, $P_{i,j}^{x+1}$ is the jth element of P_i that indicates the position of ith monarch butterfly at $(x+1)$th generation, $P_{i,j}^{x}$ is the jth element of P_i that indicates the position of ith monarch butterfly at xth generation, β is the weighting factor, and dP_j is the walk step of the ith monarch butterfly.

Similarly, the position update equation of the MS algorithm with the assumption that the moths fly only in straight line is expressed as,

$$P_{i,j}^{x+1} = \alpha\left(P_{i,j}^{x} + \eta\left(P_{best}^{x} - P_{i,j}^{x}\right)\right) \tag{13}$$

where, $P_{i,j}^{x+1}$ is the position of the ith moth at $(x+1)$th generation, $P_{i,j}^{x}$ is the position of the ith moth at xth generation, P_{best}^{x} is the position of the best moth at xth generation, α is the scale factor, and η is the acceleration factor. Substitute Eq. (12) in Eq. (13),

$$P_{i,j}^{x+1} = \alpha\left(P_{i,j}^{x+1} - \beta\left(dP_j - 0.5\right) + \eta\left(P_{best}^{x} - P_{i,j}^{x}\right)\right) \tag{14}$$

$$P_{i,j}^{x+1} - \alpha P_{i,j}^{x+1} = \alpha\left(-\beta\left(dP_j - 0.5\right) + \eta\left(P_{best}^{x} - P_{i,j}^{x}\right)\right) \tag{15}$$

$$(1-\alpha)\,P_{i,j}^{x+1} = \alpha\left(-\beta\left(dP_j - 0.5\right) + \eta\left(P_{best}^{x} - P_{i,j}^{x}\right)\right) \tag{16}$$

$$P_{i,j}^{x+1} = \frac{\alpha}{(1-\alpha)}\left(\eta\left(P_{best}^{x} - P_{i,j}^{x}\right) - \beta\left(dP_j - 0.5\right)\right) \tag{17}$$

The above equation is the update equation of the proposed MMO algorithm that tunes the DBN to produce the optimal weights.

iv) *Evaluation of population:* The population of the moths are evaluated based on the newly updated position at $(x+1)$th iteration.

v) *Termination:* The stopping criterion is performed with the production of the best solution. The pseudocode for the proposed MMO is given in algorithm 1.

Algorithm 1: Pseudocode of proposed MMO

Position estimation using MMO
Input: Solution vector $\rightarrow P_{i,j}^{x}$
Output: Best solution $\rightarrow P_{best}^{x}$
Start
Initialization of parameters
Initialize the population in random
Initialize the generation number $x = 1$,
Index γ
Acceleration factor η
maximum generation $J_{\max}$
maximum walkstep $L_{\max}$
While $\left(x < J_{\max}\right)$
For straight fly moths $\left(rand > 0.5\right)$
Update the position using equation (17)
End for
End While
Evaluate the population for updated position
Increment the generation $\left(x = x+1\right)$
Produce the best solution
Stop

3.3.b *Structure of Deep Belief Neural Network:* The weights of DBN are calculated using the proposed MMO algorithm, and the structure of DBN is shown in figure 2. The structure of DBN consists of one Multi Layer Perceptron (MLP) layer and more than one Restricted Boltzmann Machines (RBMs). Each layer of RBM and MLP indicates that the architecture of NN and the layers are

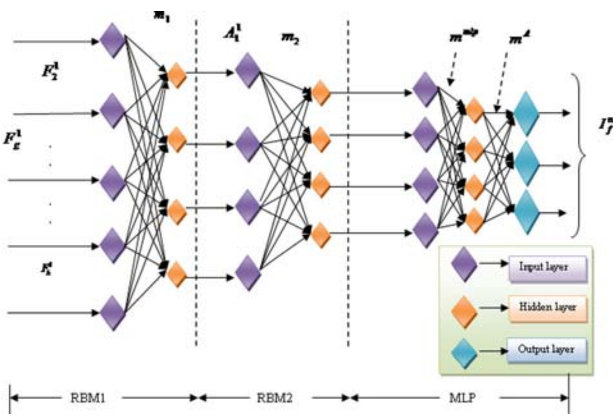

Figure 2. Structure of Deep Belief Network.

developed with the neuron interconnection. In DBN, two RBMs are considered and the input to the RBM1 is the feature vector corresponding to the extracted features. The inputs are multiplied with the input neuron weights to produce the output of the hidden layer that is fed to the RBM2 as its input. The inputs of RBM2 are processed with the hidden weights of RBM2 to produce the input of MLP layer that is processed with the weights to generate the final output.

Consider there are two RBMs, such as RBM1 and RBM2, and the input to RBM1 is the feature vector M extracted from the speech signal. The input and hidden neurons in the input layer of RBM1 are expressed as,

$$F^1 = \{F_1^1, F_2^1, F_3^1, \ldots, F_g^1, \ldots, F_k^1\}; \quad 1 \le g \le k \tag{18}$$

$$A^1 = \{A_1^1, A_2^1, \ldots, A_a^1, \ldots, A_b^1\}; \; 1 \le a \le b \tag{19}$$

where, F_g^1 is the gth input neuron that is present in the RBM1 and the total input neurons of RBM1 are equal to feature vector dimension. There are k neurons in the input layer of RBM1 to carry out the task of classification. Let the total hidden neurons in the RBM1 be, b and let ath hidden neuron in RBM2 be A_a^1. The biases of visible and hidden neurons, respectively of RBM1 are given as,

$$G^1 = \{G_1^1, G_2^1, , \ldots, G_g^1, \ldots, G_k^1\} \tag{20}$$

$$Q^1 = \{Q_1^1, Q_2^1, \ldots, Q_a^1, \ldots, Q_b^1\} \tag{21}$$

The biases of hidden and input layer of RBM1 are equal to that of the total neurons in both the layers and the weights of RBM1 are given as,

$$m^1 = \left\{m_{go}^1\right\}; \; 1 \le m \le g; \; 1 \le o \le b \tag{22}$$

where, m_{go}^1 is the weights of RBM1 and it is the weight linking the gth input neuron and ath hidden neuron of RBM1. The weight dimension is represented as $(k \times b)$. Hence, the output of RBM1 is given as,

$$A_a^1 = \tau \left[Q_a^1 + \sum_g W_g \, m_{go}^1 \right] \tag{23}$$

where, τ is the activation function in RBM1 and W_g is the feature vector as in Eq. (8). The output of the RBM1 is given as,

$$A^1 = \{A_a^1\}; \quad 1 \le a \le b \tag{24}$$

The output from RBM1 is fed as the input of RBM2 and the output of RBM2 is calculated using the equations given above. The output of RBM2 is represented as A_a^2, which is fed as the input of the MLP layer. The input neurons in MLP are given as,

$$T^n = \{T_1^n, T_2^n, \ldots, T_a^n, \ldots, T_b^n\} = \{Y_a^2\}; \; 1 \le a \le b \tag{25}$$

where, b indicates the total input neurons in MLP layer. The hidden neurons of MLP are expressed as,

$$O^n = \{O_1^n, O_2^n, \ldots, O_t^n, \ldots, O_z^n\}; \quad 1 \le y \le z \tag{26}$$

where, z is the total hidden neurons of the MLP. The bias of the hidden neurons is given as,

$$I^n = \{I_1^n, I_2^n, \ldots, I_q^n, \ldots, I_t^n\}; \; 1 \le q \le t \tag{27}$$

where, t is the output neurons in MLP layer. The weights among input and hidden layers are expressed as,

$$m^{mlp} = \{m_{oy}^{mlp}\}; \; 1 \le o \le b; \; 1 \le y \le z \tag{28}$$

where, m_{oy}^{mlp} is the weight vector between oth input neuron and the yth hidden neuron. The output of hidden layer in MLP relies on bias and weights and is given as,

$$X = \left[\sum_{a=1}^{b} m_{oy}^{mlp} \times f_a \right] m_y^n \quad \forall f_a = A_a^2 \tag{29}$$

where, m_y^n is the bias of output layer. The weight vector among the hidden and output layers is represented as m^A and is expressed as,

$$m^A = \{m_{yf}^A\}; \; 1 \le y \le z; \; 1 \le f \le t \tag{30}$$

Hence, the output of MLP is expressed as,

$$I_f = \sum_{y=1}^{z} m_{yq}^A \times X \tag{31}$$

where, m_{yq}^A is the weights among the hidden and output neurons in MLP, and X indicates the output of the hidden layer.

i) *Training of RBM layers:* The training of RBM1 and RBM2 is performed based on the unsupervised learning algorithm.
ii) *Training of MLP layer:* The MLP layer is trained using the proposed MMO algorithm, which is listed under the

supervised learning and hence, the best solution is declared based on the maximal fitness measure. The steps in the training phase of MLP layer are discussed as follows:

a) Generate the weight vectors m^A and m^{mlp} in random manner as given in Eqs. (30) and (28), respectively.
b) Read the input vector A_a^2 that is produced from output layer of RBM2.
c) Estimate the values of X and I_f with the Eqs. (29) and (31), respectively.
d) Estimate the error of MLP layer using the target and estimated output as expressed below,

$$\sigma_{avg}^1 = \frac{1}{\vartheta} \sum_{Z=1}^{\vartheta} \left(I_f - \omega \right)^2 \tag{32}$$

where, I_f indicates the attained output, ω represents the expected output, and ϑ indicates the total training samples.

e) The weights of MLP layer are updated based on the weight obtained with the proposed MMO algorithm using Eq. (17). The proposed MMO algorithm derives the optimal weights for the detection of deception.

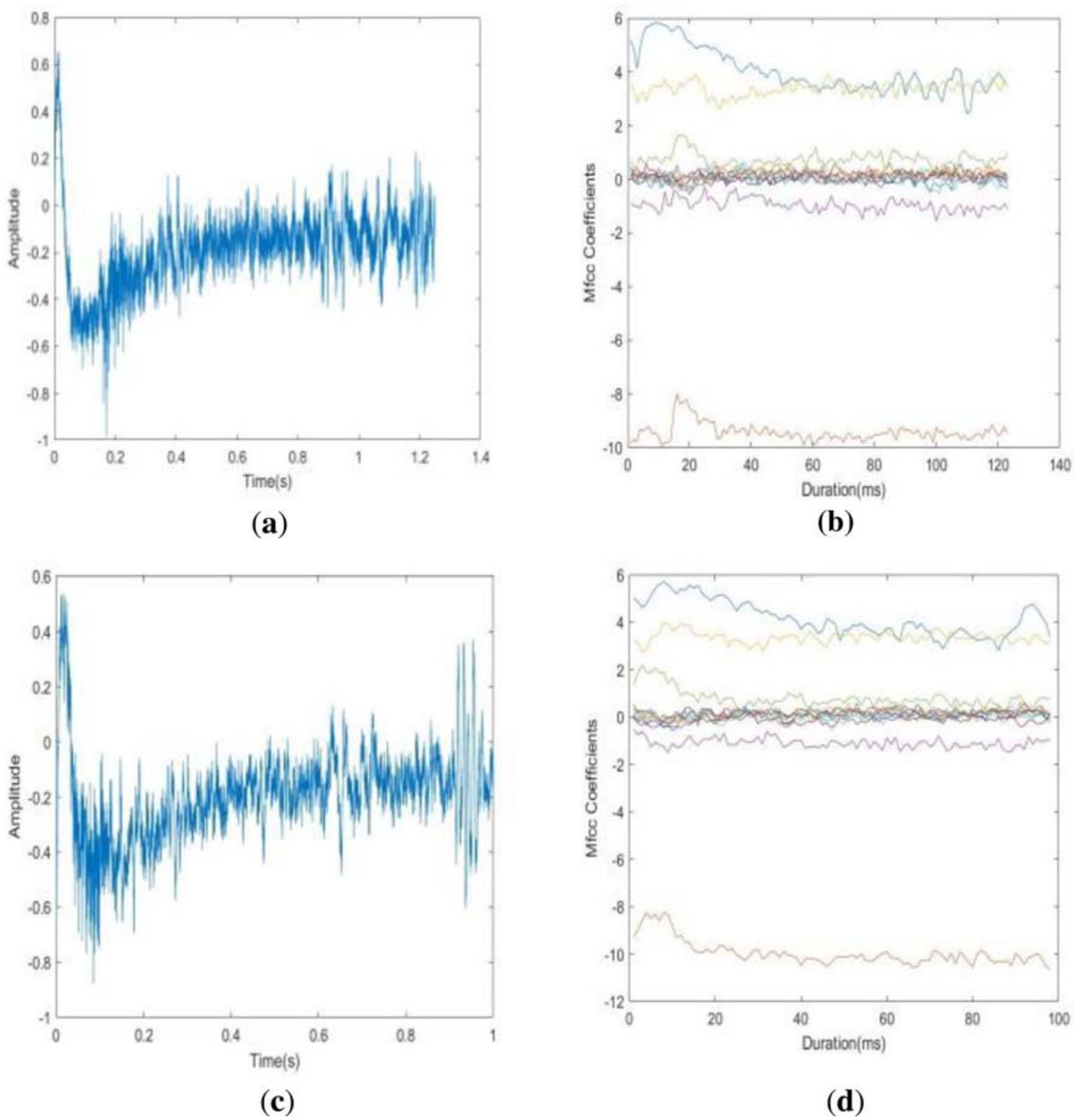

Figure 3. Sample experimental results for deception detection with the speech signal (**a**) signal 1 on pronouncing the word 'Abhishek', (**b**) extracted MFCC coefficient from signal 1, (**c**) signal 2 on pronouncing the word 'Amit', (**d**) extracted MFCC coefficient from signal 2.

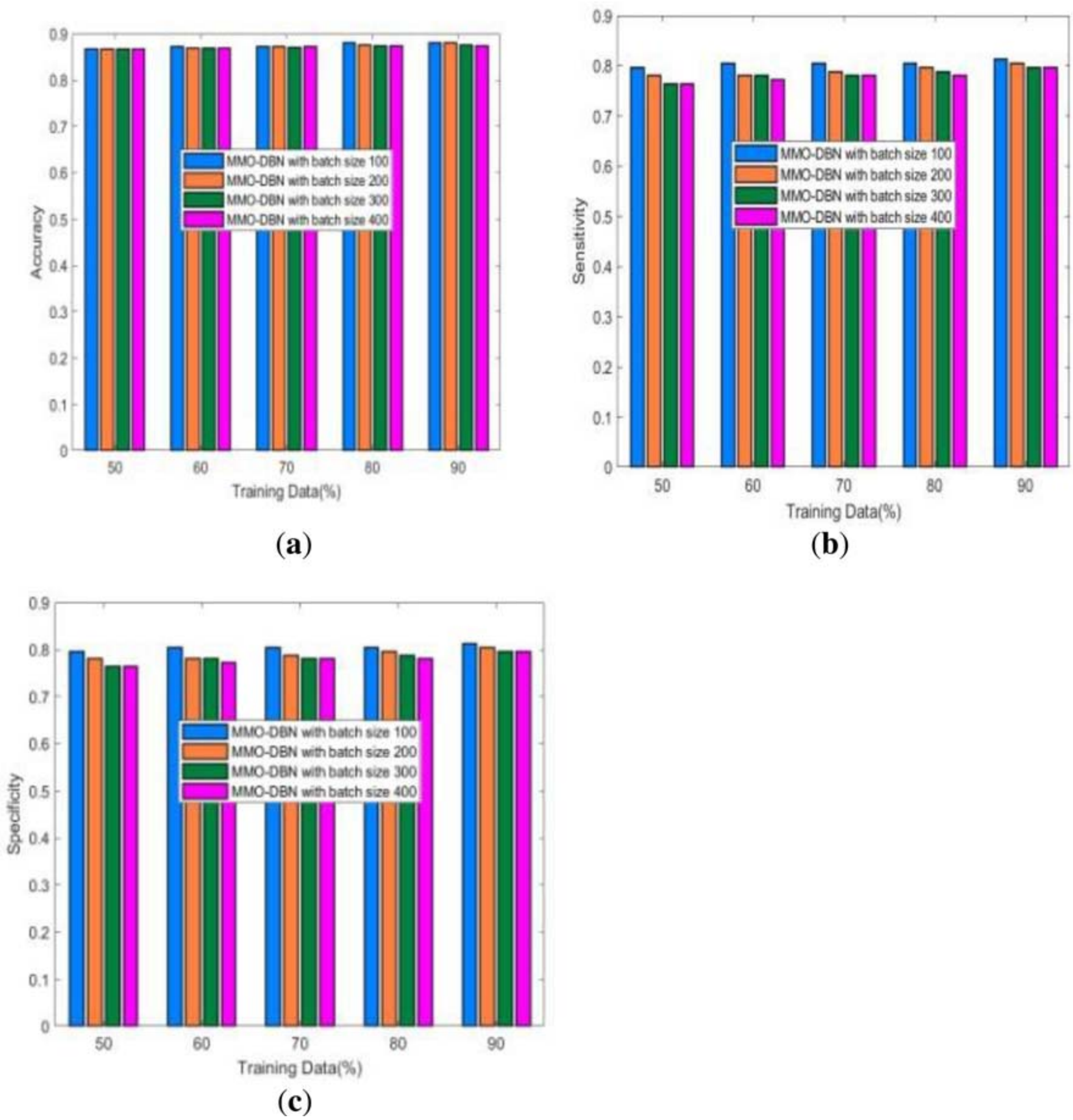

Figure 4. Performance analysis with variation in batch size: (**a**) accuracy, (**b**) sensitivity, (**c**) specificity.

f) Calculate the average error function σ^1_{avg} with the weight vector, which is updated using the proposed MMO algorithm.
g) Repeat steps b to f, until producing the best weight vector.

The classified output is the output from DBN that provides the accurate deception detection output using the speech signal of individuals under suspicion.

4. Results and discussion

The results of MMO-DBN classifier are analyzed in this section. The results of the MMO-DBN classifier when compared to the existing methods of deception detection in terms of accuracy, sensitivity and specificity are discussed.

4.1 *Experimental setup*

The classifier is implemented in MATLAB with the PC installed with Intel(R) i3 processor and Windows 10 OS using 4 GB RAM and 64-bit operating system.

4.2 *Experimental results*

This section details the experimental results of deception detection with the speech signal obtained from number of

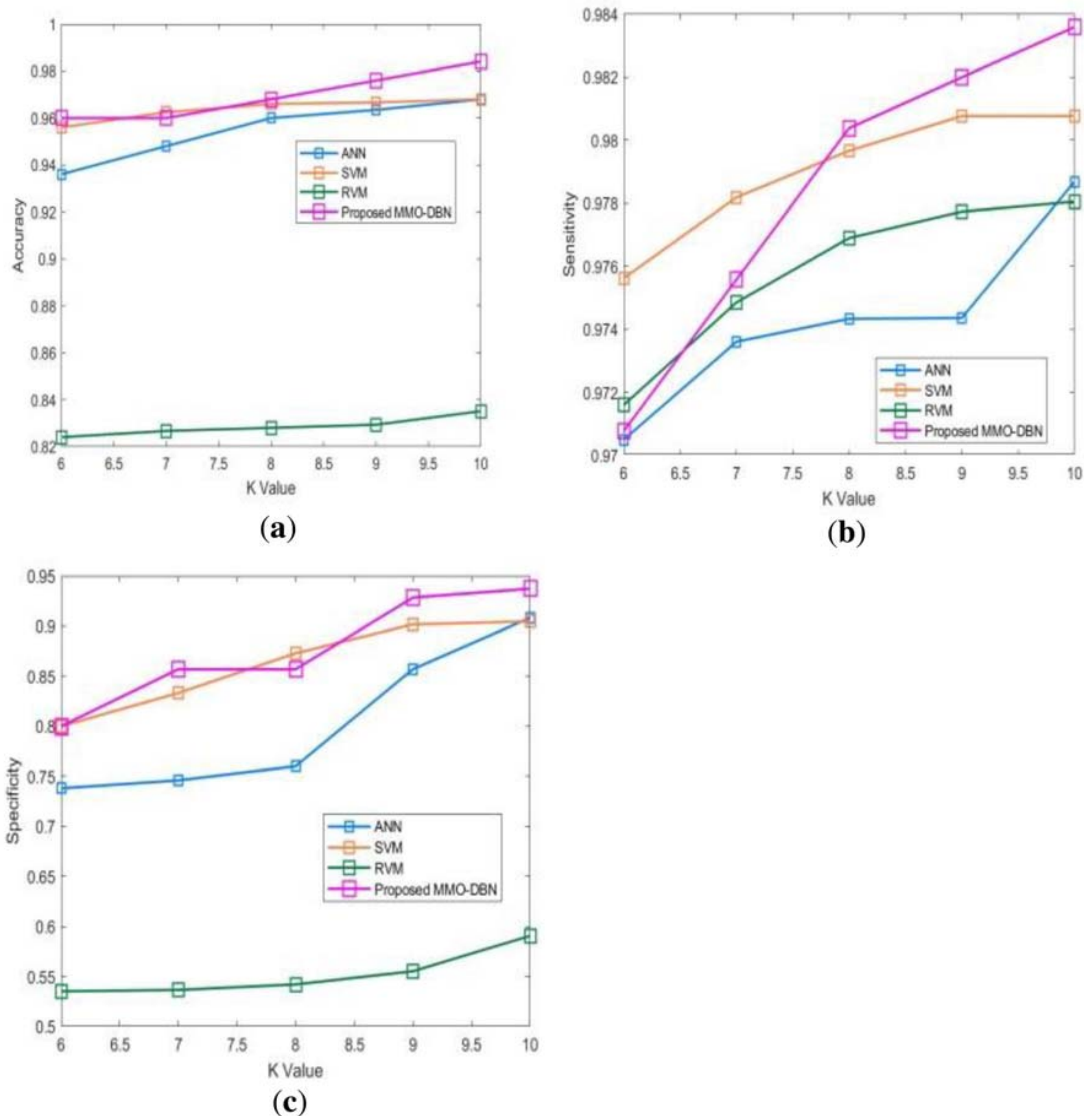

Figure 5. Comparative analysis based on kfold (**a**) accuracy, (**b**) sensitivity, and (**c**) specificity.

persons. The sample results are depicted in figure 3. Figure 3a depicts the signal corresponding to the word abhishek, figure 3b shows the extracted MFCC coefficient from the signal on pronouncement of the word abhishek, figure 3c shows the signal corresponding to the word amit, and figure 3d shows the extracted MFCC coefficient from the signal on pronouncement of the word amit.

4.3 *Evaluation metrics*

The performance of MMO-DBN algorithm is estimated in terms of the evaluation metrics, namely accuracy, sensitivity and specificity.

4.3.a *Accuracy*: The result that confirms the level of exactness is known as accuracy and is expressed as

$$\text{Accuracy} = \frac{True\,positive + True\,negative}{True\,positive + True\,negative + false\,positive + false\,negative} \tag{33}$$

4.3.b *Sensitivity:* The sensitivity is termed as the count of positives that are identified correctly.

$$TPR = \frac{True\,positive}{True\,positive + False\,negative} \tag{34}$$

4.3.c *Specificity:* The specificity is termed as the count of negatives that are identified correctly.

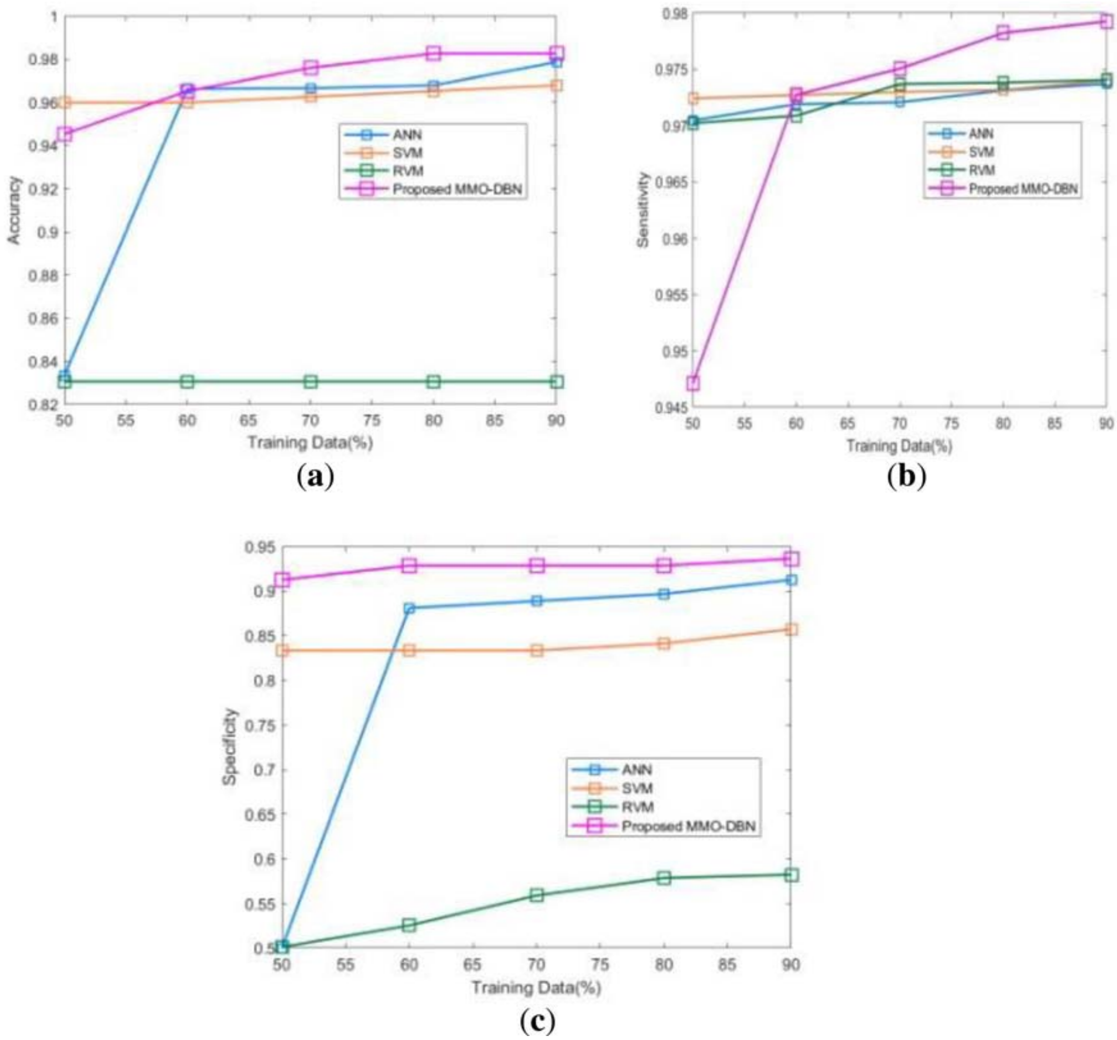

Figure 6. Comparative analysis with variation in training percentage: (**a**) accuracy, (**b**) sensitivity, (**c**) specificity.

$$TNR = \frac{True\,negative}{True\,negative + False\,positive} \tag{35}$$

4.4 *Performance analysis of MMO-DBN method in deception detection*

The performance analysis of proposed MMO-DBN method in deception detection based on the performance metrics is shown in figure 4. Figure 4a depicts the performance analysis with respect to accuracy. The batch sizes referred to the analysis include 100, 200, 300, and 400. With 90% of training, the accuracy of MMO-DBN method based on the batch sizes 100, 200, 300, and 400 is 0.8813, 0.88, 0.876, and 0.8747, respectively. Similarly, figure 4b shows the performance analysis with respect to sensitivity. With 90% of training, the sensitivity of MMO-DBN method based on the batch sizes 100, 200, 300, and 400 is 0.8127, 0.8048, 0.7968, and 0.7968, respectively. Figure 4 c shows the performance analysis with respect to specificity. With 80% of training, the specificity of MMO-DBN method based on the batch sizes 100, 200, 300, and 400 is 0.8048, 0.7968, 0.7889, and 0.781, respectively. Thus, the sensitivity, accuracy, and specificity of MMO-DBN method increase with increased percentage of training.

4.5 *Comparative methods of deception detection systems*

In this section, various existing conventional methods, such as artificial neural network (ANN) [1], relevance vector machine (RVM) [30], support vector machine (SVM) [1], are compared to the proposed MMO-DBN classifier in terms of the evaluation metrics, namely sensitivity, accuracy, and specificity.

Table 1. Comparative discussion involving the techniques of deception detection.

Variation in values	Metrics	Methods			
		ANN	SVN	RVM	Proposed MMO-DBN
k-fold	Accuracy	0.968	0.968	0.8351	**0.984**
	Sensitivity	0.9787	0.9808	0.9781	**0.9836**
	Specificity	0.9085	0.9048	0.5907	**0.9375**
Training percentage	Accuracy	0.9787	0.968	0.8307	**0.9827**
	Sensitivity	0.9737	0.974	0.9741	**0.9793**
	Specificity	0.9127	0.8571	0.5823	**0.9365**

The bold values represent the best performance

4.6 *Comparative analysis of the methods of deception detection systems*

The comparative analysis of each method is detailed and the analysis of the method is performed with the variation in k-fold value and the training percentages.

4.6.a *Comparative analysis based on k-fold value:* The analysis with the variation in kfold value for comparing the effectiveness of the MMO-DBN over the existing methods is depicted in figure 5. In figure 5a the accuracy of methods on the basis of change in kfold is shown. When the k-fold is 10, the accuracy obtained by the ANN is 0.968, SVM is 0.968, RVM is 0.8351, and the proposed MMO-DBN is 0.984. Figure 5b shows sensitivity of the comparative techniques based on the variation in kfold. When the k-fold is 10, the sensitivity produced by the ANN is 0.9787, SVM is 0.9808, RVM is 0.9781, and the proposed MMO-DBN is 0.9836. Figure 5c shows specificity of techniques with the variation in kfold. When the k-fold is 10, the specificity produced by the ANN is 0.9085, SVM is 0.9048, RVM is 0.5907, and the proposed MMO-DBN is 0.9375.

4.6.b *Comparative analysis based on training percentage:* The analysis on the basis of training percentage to compare the effectiveness of the proposed method over the existing methods is shown in figure 6. In figure 6a, the accuracy of the comparative models with variation in percentage of training is shown. With 90% of training, the accuracy of ANN is 0.9787, SVM is 0.968, RVM is 0.8307, and the proposed MMO-DBN is 0.9827. Figure 6b shows the sensitivity of methods with varying percentage of training. With 90% of training, the sensitivity of ANN is 0.9737, SVM is 0.974, RVM is 0.9741, and the proposed MMO-DBN is 0.9793. Figure 6c shows specificity of the comparative methods with the variation in training percentage. With 50% of training, the specificity of ANN, SVM, RVM, and the proposed MMO-DBN is 0.5, 0.8333, 0.5012, and 0.9127, respectively.

4.7 *Comparative discussion*

Table 1 depicts the comparison of MMO-DBN over the existing models, such as ANN, SVM, and RVM to show the better performance of MMO-DBN method. When the existing ANN, SVM, and RVM have the accuracy of 0.968, 0.968, and 0.8351, respectively, the MMO-DBN obtained an accuracy of 0.984 with respect to the k-fold. Similarly, the sensitivity of MMO-DBN method and the existing ANN, SVM, and RVM with respect to the k-fold is 0.9808, 0.9781, 0.9836, and 0.9787, respectively. In the same way, the specificity of the methods, namely ANN, SVM, RVM, and the proposed MMO-DBN with respect to the k-fold is 0.9085, 0.9048, 0.5907, and 0.9375, respectively. When the existing ANN, SVM, and RVM have the accuracy of 0.9787, 0.968, and 0.8307, respectively, the accuracy of the MMO-DBN method is 0.9827 in terms of training percentage. Similarly, the sensitivity of MMO-DBN method, and the existing ANN, SVM, and RVM in terms of the training percentage is 0.9793, 0.9737, 0.974, and 0.9741, respectively. In the same way, the specificity of the methods, namely ANN, SVM, RVM, and the proposed MMO-DBN in terms of training percentage is 0.9127, 0.8571, 0.5823, and 0.9365, respectively. Hence, the proposed MMO-DBN method offers increased accuracy, sensitivity, and specificity as compared to other methods.

5. Conclusion

Deception detection is an extremely tedious task, which cannot be performed with high reliability and accuracy manually. Thus, this paper proposes an automatic deception detector technique called Moth Monarch optimization-based Deep Belief Neural Network (MMO-DBN) for the detection of deception with increased accuracy. The proposed MMO-DBN method comprises two phases, namely feature extraction and classification. The noise signals present in the input signal are removed with the pre-processing step, and the enhanced signal is fed to feature extraction to extract the necessary features. Then, these features are fed to classification process using Deep Belief Neural Network (DBN) that is trained using the proposed Moth Monarch optimization (MMO) algorithm, which is the integration of Monarch Butterfly Optimization (MBO) and Moth Search (MS) algorithm. The analysis of the proposed MMO-DBN is performed with the evaluation metrics, such as sensitivity, accuracy, and specificity. The MMO-DBN method obtained an increased accuracy of 0.984, sensitivity of 0.9836, and specificity of 0.9375, which is high as compared to the existing techniques. In future, this method will be enhanced to deal with other problems related to the classification of gender and regional languages.

References

[1] Srivastava N and Dubey S 2018 Deception detection using artificial neural network and support vector machine. In: *Proceedings of Second International Conference on Electronics, Communication and Aerospace Technology*, pp. 1205–1208

[2] Zhou L, Twitchell D P, Qin T, Burgoon J K and Nunamaker J F 2003 An exploratory study into deception detection in text-based computer-mediated communication. In: *Proceedings of the International Conference on System Sciences,* p. 10

[3] Abouelenien M, Pérez-Rosas V, Mihalcea R and Burzo M 2014 Deception detection using a multimodal approach. In: *Proceedings of the International Conference on Multimodal Interaction*, pp. 58–65

[4] Perez-Rosas V, Mihalcea R, Narvaez A and Burzo M 2014 A multimodal dataset for deception detection. In*: LREC*, pp. 3118–3122

[5] Tsiamyrtzis P, Dowdall J, Shastri D, Pavlidis I, Frank M G and Ekman P 2005 Lie detection-recovery of the periorbital signal through tandem tracking and noise suppression in thermal facial video. In: *Proceedings of SPIE Sensors, and Command, Control, Communications, and Intelligence Technologies for Homeland Security and Homeland Defense,* vol. 5778, pp. 29–31

[6] Barone V, Ferrera G and Romeo N 2002 A clinical case report of priapism treated with external RT. *Tumori J.* 88(6): 538–539.

[7] Ahsan A 2014 Crime detection using voice behaviour on the basis of collected evidence, *Int. J. Eng. Adv.Technol* 4(1): 137–141.

[8] Burkhardt F, Paeschke A, Rolfes M, Sendlmeier W F and Weiss B 2005 A database of german emotional speech. *Interspeech* 5: 1517–1520.

[9] Feng Y H and Wang G G 2018 Binary Moth search algorithm for discounted {0-1} Knapsack problem. *IEEE Access* 6: 10708–10719.

[10] Sun L, Chen S, Xu J and Tian Y 2019 Improved Monarch butterfly optimization algorithm based on opposition-based learning and random local perturbation, *Complexity* 4: 1–20.

[11] Chow A and Louie J 2005 Detecting lies via speech patterns. 7.

[12] Cristin R, Cyril Raj V and Marimuthu R 2019 Face image forgery detection by weight optimized neural network model. *Multimed. Res. (MR)* 2(2): 19–27.

[13] An G, Levitan S I, Hirschberg J and Levitan R 2018 Deep personality recognition for deception detection. In: *Proceedings of Inter speech*, pp. 421–425

[14] Arul V H, Sivakumar V G, Marimuthu R and Chakraborty B 2019 An approach for speech enhancement using deep convolutional neural network. *Multimed. Res. (MR)* 2(1): 37–44.

[15] Menaga D and Revathi S 2020 Deep learning: a recent computing platform for multimedia information retrieval. In: *Deep Learning Techniques and Optimization Strategies in Big Data Analytics,* pp. 124–141.

[16] Hirschberg J, Benus S, Brenier J M, Enos F, Friedman S, Gilman S, Girand C, Graciarena M, Kathol A, Michaelis L and Pellom B L 2005 Distinguishing deceptive from non-deceptive speech. In*: Ninth European Conference on Speech Communication and Technology*

[17] Krishnamoorthy N and Revathi K 2015 Evaluating the Performance of Swallow Swarm Optimization Algorithm. *Int. J. Appl. Eng. Res.* 10(20): 17021–17025.

[18] Waykar S B and Bharathi C R 2019 Intent aware optimization for content based lecture video retrieval using Grey Wolf optimizer. *J. Eng. Res.* 7(3): 123–141.

[19] Pérez-Rosas V and Mihalcea R 2014 Cross-cultural deception detection. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 440–445

[20] Lai Y F, Chen M Y and Chiang H S 2018 constructing the lie detection system with fuzzy reasoning approach, *Granul. Comput.* 3(2): 169–176.

[21] Abootalebi V, Moradi M H and Khalilzadeh M A 2009 A new approach for EEG feature extraction in P300-based lie detection, *Comput. Methods Prog. Biomed.* 94(1): 48–57.

[22] Chen S M and Chung N Y 2006 Forecasting enrollments of students by using fuzzy time series and genetic algorithms. *Int. J. Inf. Manag. Sci.* 17(3):1–17.

[23] Hernández-Castañeda Á, Calvo H, Gelbukh A and Flores J J G 2017 Cross-domain deception detection using support vector networks, *Soft Comput.* 21(3): 585–595.

[24] Pérez-Rosas V, Abouelenien M, Mihalcea R and Burzo M 2015 Deception detection using real-life trial data. In: *Proceedings International Conference on Multimodal Interaction*, pp. 59–66.

[25] Pérez-Rosas V, Abouelenien M, Mihalcea R, Xiao Y, Linton C J and Burzo M 2015 Verbal and nonverbal clues for real-life deception detection. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* pp. 2336–2346.

[26] Wang G G, Deb S, Zhao X and Cui Z 2018 A new monarch butterfly optimization with an improved crossover operator, *Oper. Res.* 18(3): 731–755.

[27] Wang G G, Deb S and Cui Z 2019 Monarch butterfly optimization, *Neural Comput. Appl.* 31(7): 1995–2014.

[28] Wang G G 2018 Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems, *Memet. Comput.* 10(2): 151–164.

[29] Gogate M, Adeel A and Hussain A 2017 Deep learning driven multimodal fusion for automated deception detection. *Comput. Intell.* 1–6

[30] Zhou Y, Zhao H, Pan X and Shang L 2015 Deception detecting from speech signal using relevance vector machine and non-linear dynamics features. *Neurocomputing* 151: 1042–1052.

[31] Abouelenien M, Pérez-Rosas V, Mihalcea R and Burzo M 2017 Detecting deceptive behavior via integration of

discriminative features from multiple modalities. *IEEE Trans. Inf. Forensics Secur.* 12(5): 1042–1055.

[32] O'Shea J, Crockett K, Khan W, Kindynis P, Antoniades A and Boultadakis G 2018 Intelligent Deception Detection through Machine Based Interviewing. In: *Proceedings of International Joint Conference on Neural Networks*, pp. 1–8.

[33] Nasri H, Ouarda W and Alimi A M 2016 ReLiDSS: Novel lie detection system from speech signal. In: *Computer Systems and Applications,* pp. 1–8.

[34] OrShea J, Crockett K, Khan W and Kindynis P 2018 Intelligent deception detection through machine based interviewing. In: *International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil*.

[35] Twitchell D P and Fuller C M 2019 Advancing the assessment of automated deception detection systems: incorporating base rate and cost into system evaluation, *Inf. Syst. J.* 29(3): 738–761.

[36] Ali M Z M, Mansor W, Khuan L Y and Zabidi A 2012 Simulink model of Mel frequency cepstral coefficient analysis for extracting asphyxiated infant cry features. In: *Proceedings of International Conference on Biomedical Engineering*.

[37] Tzanetakis G and Cook P 2002 Musical genre classification of audio signals, *IEEE Trans. Speech Audio Process.* 10(5): 293–302.

[38] Vrabie V, Granjon P and Serviere C 2003 Spectral kurtosis: from definition to application. In: *6th IEEE International Workshop on Nonlinear Signal and Image Processing (NSIP 2003)*.

[39] Peeters G, Giordano B L, Adams S M Susini P and Misdariis N 2011 The Timbre Toolbox: extracting audio descriptors from musical signals. *J. Acoust. Soc. Am.* 130(5): 2902–2916.

[40] Pavlidis I and Levine J 2001 Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph. In: *Proceedings of 23rd Annual EMBS International Conference*.

[41] Vojt B J 2016 Deep neural networks and their implementation. *Department of Theoretical Computer Science and Mathematical Logic*