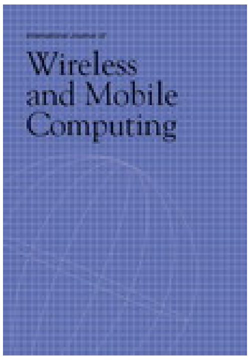


International Journal of Wireless and Mobile Computing

ISSN online: 1741-1092 - ISSN print: 1741-1084
https://www.inderscience.com/ijwmc

Metaheuristic-assisted deep ensemble technique for identifying sarcasm from social media data

Geeta Abakash Sahu, Manoj Hudnurkar

DOI: 10.1504/IJWMC.2024.10062122

Article History:

Received:	19 April 2022
Last revised:	22 February 2023
Accepted:	22 February 2023
Published online:	07 February 2024

Copyright © 2024 Inderscience Enterprises Ltd.

Metaheuristic-assisted deep ensemble technique for identifying sarcasm from social media data

Geeta Abakash Sahu*

Faculty of Computer Studies (FoCS),
Symbiosis International University (SIU),
Symbiosis Centre for Research & Innovation (SCRI),
Pune, India
Email: geetasahu23@gmail.com
*Corresponding author

Manoj Hudnurkar

Operations and IT,
Symbiosis International University (SIU),
Symbiosis Centre for Management &
Human Resource Development (SCMHRD),
Pune, India
Email: manojh786@outlook.com

Abstract: Sarcasm is regarded as the enveloping linguistic factor in online documents that describes the deeply-felt subjective and opinions. This paper intends to introduce a sarcasm detection model that classifies words under sarcastic or non-sarcastic forms. Pre-processing is the initial phase, where the stop word removal and tokenisation are performed. The pre-processed data is then subjected to extracting the features, where, 'information gain, chi-square, mutual information, and symmetrical uncertainty-based features' are extracted. As the curse of dimensionality becomes the greatest crisis, optimal feature selection is carried out. For sarcasm detection, an ensemble classifier such as NN, RF, SVM and optimised DCNN is used, in which the weights of DCNN are optimally selected. For optimal feature selection and optimised DCNN, a hybrid optimisation model termed as Clan Updated Grey Wolf Optimisation (CU-GWO) is proposed. Finally, the effectiveness of the proposed algorithm is compared with extant methods in terms of various measures.

Keywords: sarcasm; tokenisation; information gain; optimised DCNN; CU-GWO optimisation.

Reference to this paper should be made as follows: Sahu, G.A. and Hudnurkar, M. (2024) 'Metaheuristic-assisted deep ensemble technique for identifying sarcasm from social media data', *Int. J. Wireless and Mobile Computing*, Vol. 26, No. 1, pp.25–38.

Biographical notes: Geeta Abakash Sahu is a research scholar, pursuing a PhD from Symbiosis Center for Research and Innovation (SCRI), Symbiosis International (Deemed University), Pune, Maharashtra, India. Her research interests include machine learning and artificial intelligence. She has published a few papers in international journals of repute on machine learning. She has got 14 years of teaching experience and currently an Assistant Professor at Vidyalankar School of Information Technology (VSIT), Mumbai, Maharashtra, India.

Manoj Hudnurkar is a Professor and Head, in Operations and Analytics at Symbiosis Centre for Management and Human Resource Development (SCMHRD), Symbiosis International (Deemed University), Pune, Maharashtra, India. Previously, he worked in the industry as a Project Manager and has handled various projects. He has published various papers in international journals of repute on collaborative supply chain management, supplier classification, and supplier performance management. He also has experience in consulting projects and corporate training.

1 Introduction

Sarcasm is defined as a specific type of sentiment through which people express their negative feelings via positive or intensified positive words in the text. Also, this is said to be a complex linguistic phenomenon to express scorn or ridicule. In sarcasm expression, actual meaning is often the contrary of factual meaning, which makes sarcasm multifaceted for processing languages (Ludlow et al., 2017; Sulis et al., 2016; Deliens et al., 2017; Matsui et al., 2016). In the meantime, sarcasm is also a general way to convey emotions in everyday life. It is frequently deployed for expressing nasty feelings or

Copyright © 2024 Inderscience Enterprises Ltd.

making a fun with someone. As sarcastic expression is extensively utilised in social media, daily communications and product reviews, detection of sarcasm has attracted the attention of the NLP analysts, intending to recognise sarcastic expression from texts automatically (Schouten and Frasincar, 2016; Ebrahimi et al., 2017). Sarcasm detection turned out to be a demanding NLP task, and has been broadly exploited in human-machine dialogues, sentimental analysis and other NLP appliances (Lagerwerf, 2007; Creusere, 1999).

Usually, sarcasm expression involves two key factors: '(a) the contrast of sentiment polarity in the sentence and (b) the contrast between the conveyed sentiment and the situation'. Certain analysts have utilised sentiment data to identify sarcasm depending on feature engineering (Sanders, 2013; Uchiyama et al., 2006; Persicke et al., 2013). Most of the previous sarcasm recognition approaches are based on manually formulated sentimental features. Numerous analysts have extracted features from sentimental data to detect sarcasm with conventional machine learning schemes. As deriving features entails a large amount of physical labour, certain studies have attempted to make use of deep learning to solve the issue in current years (Justo et al., 2014; Del Pilar Salas-Zárate et al., 2017; Ashwitha et al., 2020).

Techniques depending on deep learning schemes have an attempted reputation in NLP appliances as they are able to handle data sparseness in unbalanced data sets. Analysis on verbal aggression and bullying has been reported by DNN for sentimental analysis on short-texts. Moreover, a lot of techniques like SVM, RF and CNN have been deployed for automated sarcasm detection (Sonawane and Kolhe, 2020; Kunneman et al., 2015; Beno et al., 2014). However, precise outcomes were not attained using these approaches, and therefore, a sophisticated model is necessary to be developed for more satisfactory outputs. Moreover, many optimisation algorithms are utilised to solve optimisation issues in many fields (Cai et al., 2021; Cui et al., 2019; Zhao et al., 2022). Thus, we utilise a novel hybrid optimisation technique for sarcasm detection.

The contribution of the work is given below:

- Extracting text-based features like information gain, chi-square, mutual information and symmetrical uncertainty-based features, and choosing optimal features from the extracted feature set for precise detection.
- To detect the sarcasm in the text, ensemble classifiers like RF, SVM, NN and DCNN are utilised. Also, the weights of the DCNN are fine-tuned for precise detection.
- For optimal features and optimal weights selection, a novel Clan Updated Grey Wolf Optimisation (CU-GWO) model is proposed, which is the combined version of GWO and EHO algorithms.

Here, Section 2 reviews the work. Section 3 portrays proposed sarcasm detection model and its implications. Section 4 discusses the obtained results. Section 5 concludes the paper.

2 Literature review

2.1 Related works

Mukherjee and Bala (2017) attempted to offer knowledge to the scheme by taking the various sets of characteristics into consideration that were associatively text-independent, called part of speech n-gram. The various characteristics set with a range were tested by deploying the 'fuzzy clustering and NB algorithms'. The investigation outcome showed that the sarcasm detection process has attained the profit from the features inclusion that incarcerates authorial style within the microblog authors. Further, the achievement was 65% better for sarcasm detection than any other methods.

Ren et al. (2018) explored NN methods for sarcasm detection on twitter. Further, two various context-based neural techniques were proposed on the basis of CNN. The simulation outcome has demonstrated that the developed neural model has higher performance over other conventional methods. In the interim, the sarcastic clues from contextual information have been efficiently decoded by the implemented context-augmented neural methods and have explored an augmented enhancement in the performance of the detection process.

Keerthi and Harish (2018) proposed a new technique to categorise sarcastic texts by means of a content-oriented feature selection technique. The developed model comprises two stage feature selection methods for selecting most representation features. Initially, extant feature selection techniques such as MI, and information feature subset were refined by means of a k-means model for selecting the most similar features. The chosen features were categorised by means of 2 classifiers: RF and SVM.

Lu and Yang (2020) proposed a 'multi-level memory network' by means of sentimental semantics for capturing the characteristics of sarcasm expression. Here, the 1st-level memory network for capturing sentimental semantics, and employ the 2nd-level memory network for capturing the difference among sentimental semantics and the circumstances in all sentences. In addition, an enhanced CNN was used to develop the memory network in non-existence of local data.

Deepak et al. (2020) proposed deep learning in sarcasm detection by means of code-switch tweets, particularly the mash-up of Indian native languages like Hindi with English. The adopted approach was a hybridised version of 'Bi-LSTM with a softmax attention layer and CNN' for real-time detection of sarcasm. The approach outperformed the other existing schemes with better F-measure and precision rates.

Bharti et al. (2016) developed a technique depending on the Hadoop structure for taking real-time tweets and then processing them using an algorithm set, in which the sarcastic sentiment was determined effectively. From the result, it was made clear that the proposed Hadoop algorithm possessed a considerable performance that was better than any other traditional models in terms of elapsed time.

Bouazizi and Ohtsuki (2016) developed a pattern-based model for the detection of sarcasm on Twitter. Here, four feature sets were proposed to define the various kinds of sarcasms. By using this, the tweets were classified in accordance with their sarcastic and non-sarcastic nature. Hence, the implemented model has attained a better accuracy and precision of 83.1% and 91.1%. Further, the importance of every proposed set of features was learned and the added values were computed to the classification. Particularly, the importance of pattern-based features was underlined in this for sarcastic statement detection.

Son et al. (2019) presented a 'sAtt-BLSTM convent' depending on the hybrid 'sAtt-BLSTM and convNet deploying GloVe' to form semantic word embedding. Moreover, feature maps were produced by sAtt-BLSTM; and auxiliary features were combined to form convNet. The enhancement of the developed model was proved against other existing models.

2.2 Review

Table 1 reviews the traditional models in terms of sarcasm detection. Initially, NB (Mukherjee and Bala, 2017) gains increased sarcasm detection accuracy and better performance of function words. But, it was not suitable for large data sets and the classification and clustering were not well adopted. Context-augmented neural network (Ren et al., 2018) has better effectiveness and good twitter sarcasm detection. Even though, needs improvement in detection performance and lacks in dealing with more contextual tweets. SVM (Kumar and Harish, 2018) possess a better precision, recall and F-measures values and selects relevant feature subsets. However, the classification accuracy still needs improvement, and this method still needs extension. CNN (Ren et al., 2020) has better detection of sarcasm and high efficiency. In spite of this, this method lacks assistance in sentiment analysis. CNN (Jain et al., 2020) offers better accuracy and an enhanced F-measure. The major challenge of this methodology was that it relies mainly on online language identifiers. The Hadoop model (Bharti et al., 2016) has reduced processing time and better detection of sarcasm in tweets. However, needs an attainment of sufficient data sets and further deployment of various algorithms. SVM (Bouazizi and Ohtsuki, 2016) is the methodology that has better extraction of patterns and is more efficient. But, this method needs improvement in sentiment analysis performance opinion mining. The sAtt-BLSTM convNet (Son et al., 2019) has better accuracy and high precision. However, the mash up languages was not included.

Table 1 Review on conventional sarcasm detection models

Author	*Developed method*	*Features*	*Challenges*
Mukherjee and Bala (2017)	NB	• Increased sarcasm detection accuracy. • Better performance of function words.	• Not suitable for large datasets.
Ren et al. (2018)	Context-augmented neural network	• Better effectiveness. • Good twitter sarcasm detection.	• Detection performance needs improvement. • Lacks in dealing with more contextual tweets.
Kumar and Harish (2018)	SVM	• Better precision, recall and F-measures. • Selects relevant feature subsets.	• Classification accuracy still needs to improved. • Needs further extension.
Lu and Yang (2020)	CNN	• Better detection of sarcasm. • High efficiency.	• No assistance on sentiment analysis.
Deepak et al. (2020)	CNN	• Better accuracy. • Enhanced F-measure.	• Relies mainly on online language identifier.
Bharti et al. (2016)	Hadoop model	• Reduced process time. • Detect sarcasm in tweets.	• Need attainment of sufficient data sets. • Further needs deployment of various algorithms.
Bouazizi and Ohtsuki (2016)	SVM	• Better extraction of patterns. • More efficient.	• Sentiment analysis performance needs improvement. • Opinion mining needs to be enhanced.
Son et al. (2019)	sAtt-BLSTM convNet	• Better accuracy. • High precision.	• Mash up languages was not included.

3 Proposed sarcasm detection model

Figure 1 shows the illustration of the proposed sarcasm detection model. The implemented Sarcasm detection approach encompasses 5 vital phases such as: pre-processing, extraction of appropriate features, selecting the optimal features and deep learning based detection. At first, stop word removal and tokenisation are deployed during pre-processing. Then, the features including information gain (Fe_{IG}), chi-square (Fe_{CS}), mutual information (Fe_{MI}) and symmetrical uncertainty-based features (Fe_{SU}) are extracted. Since the derived features endure the problem of curse of dimensionality, selection of optimal features is mandatory. The selected optimal features are then provided as the input to an ensemble technique that includes RF, SVM, NN and DCNN for detecting the sarcasm existing in the text input. The flow of the ensemble technique is as follows: output of RF, SVM and NN is given as the input to an optimised DCNN that provides the detected output. To solve the optimisation issues (finding optimal features and weight tuning of DCNN), this work exploits a new CU-GWO model.

Furthermore, the objective function *Obj* defined to attain the precise detection results is given in equation (1), where Loss is determined in equation (14). This determines the maximisation of detection accuracy. Considering the given objective, the proposed hybrid algorithm ensures the attainment of optimal solutions.

$$Obj = \text{Min}(Loss) \tag{1}$$

3.1 Pre-processing

This is the preliminary step; where the keywords are extracted from each domain once the stop words are removed. Stop words are natural language divisions that provide better knowledge about chats, whether they are illegal or legal. Usually, the stop words are articles and pronouns that do not offer meaning to the sentence. When the stop words are recognised to be malicious, they have to be eliminated. This removal minimises the term space dimensionality. Thereby, the key words are extracted. The example for the pre-processing technique is shown in Table 2.

Table 2 Example for stop word removal and key word extraction

Sample text	*Text without stop words*
Flight has not arrived yet	Flight not arrived
Thanks for the response. We are hopeful!	Thanks, response, hope
I like reading, so I read	Like, reading, read
Can listening be exhausting?	Listening, exhausting
A computer science portal for Geeks	Computer science, portal, Geeks

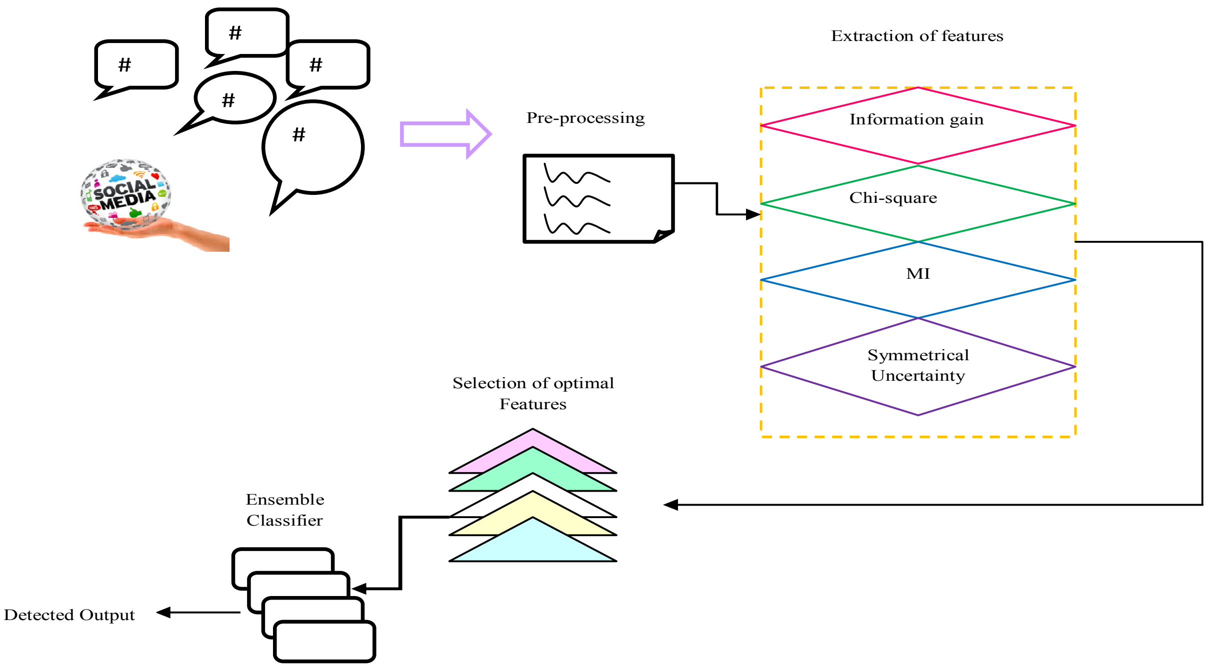
3.2 Feature extraction

3.2.1 Chi-square features

It is a renowned feature generation model. Here, the chi-square values among target feature and other features are computed (Gazeloglu, 2020). By this computation, the optimal chi-square score is found and the preferred number of properties is elected. It is formulated as shown in equation (2), wherein, *OF* refers to observed frequency and *EF* refers to expected frequency.

$$\chi^2 = \frac{(OF - EF)^2}{EF} \tag{2}$$

Figure 1 Overall architecture of the proposed sarcasm detection model (see online version for colours)



3.2.2 *MI features*

It is defined as the measure of exchanged information between two ensembles of random variables U and V (Sarhrouni et al., 2012). It is formulated as shown in equation (3), wherein, ρ signifies the probability.

$$MI = \sum \rho(U,V) \log_2 \frac{\rho(U,V)}{\rho(U)\rho(V)} \tag{3}$$

The extracted MI features are denoted by Fe_{MI} .

3.2.3 *Information gain features*

It is a significant feature that is formulated as shown in equation (4), in which IG refers to information gain and w signifies weight (Sarhrouni et al., 2012).

$$IG = H(U) - (w * H(V)) \tag{4}$$

3.2.4 *Symmetrical uncertainty-based features*

It is a normalised form of MI and is formulated as shown in equation (5), wherein, H refers to the entropy (Sarhrouni et al., 2012).

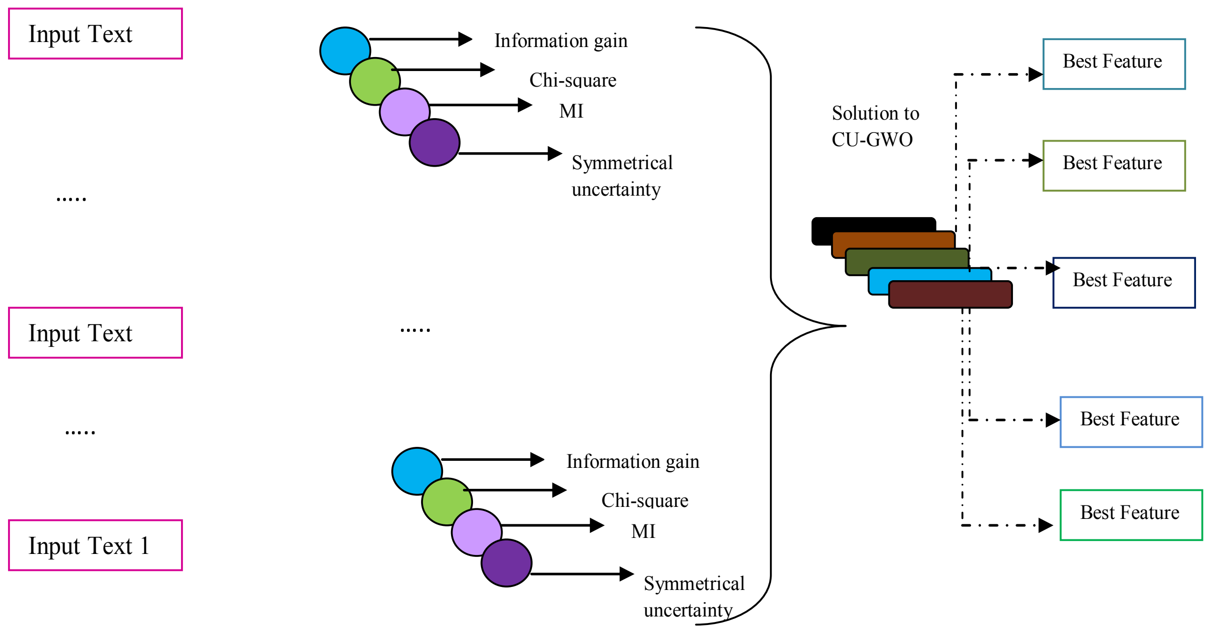
$$SU = 2.\frac{MI(U,V)}{H(U)H(V)} \tag{5}$$

Accordingly, the extracted chi-square, information gain, MI and symmetrical uncertainty features are summed up as Fe, i.e., $Fe = \chi^2 + IG + MI + SU$.

3.3 *Optimal feature selection and deep ensemble technique for sarcasm detection*

While extracting all the features mentioned above, the training become congested with precise and unwanted information. This may be termed as the issue of 'curse of dimensionality'. This may affect the performance of the model to produce accurate detection results. Hence, an attempt is made to find the optimal features for better detection results. To select the optimal features, this paper intends to introduce a new hybrid optimisation model, which enrols its selection based on the objective mentioned in this work. Figure 2 shows the illustration of optimal feature selection. For different input text, the information gain, chi-square, MI and symmetrical uncertainty features are extracted. From the extracted features, the optimal (best) features are selected with the aid of the proposed CU-GWO algorithm.

Figure 2 Optimal feature selection process (see online version for colours)



3.3.1 Ensemble technique

This paper models the ensemble technique to detect the presence of sarcasm in the given input text. There are two key reasons for choosing an ensemble classifier over a single model are:

- *Performance*: Compared to a single contributing model, an ensemble can anticipate events more accurately and perform better overall.
- *Robustness*: An ensemble narrows the prediction and model performance distribution.

The proposed ensemble concept includes classifiers like RF, SVM, NN and deep CNN. The procedure for this ensemble is as follows: The selected optimal features are directly given to the classifiers like RF, SVM and NN. The outputs from these classifiers are again given to the DCNN model, from which the final detection output is determined. Figure 3 shows the representation of the proposed ensemble technique.

RF: 'This (Entezari-Maleki et al., 2009) is a classification algorithm that combines multiple weak classifiers (decision trees) to form a stronger classifier'. Here, classification is done by allowing these trees to elect the renowned class. The features of each produced tree are randomly selected at every node. In addition, a general CART DT exploits each feature and as a result, the RF model ensures the randomness of features. ARF approach offers high classification accuracy than a DT; nevertheless the interpretability is not obvious as the features with significant role are unrecognised. The output of the RF classifier is denoted by.

SVM: SVM considers non-linear mapping to transform the original training data into a higher dimension. It explores the linear optimum hyper plane within this new dimension (Entezari-Maleki et al., 2009; Thomas and Rangachar, 2019). A hyper plane is a decision boundary separates the tuples of each class. The SVM discovers the hyper plane by means of margins and support vectors. The hyper plane of 2-class linear separable problem is evaluated as shown in equation (6). Here, hp denotes the distance between origins to hyper plane and VT indicates the normal vector.

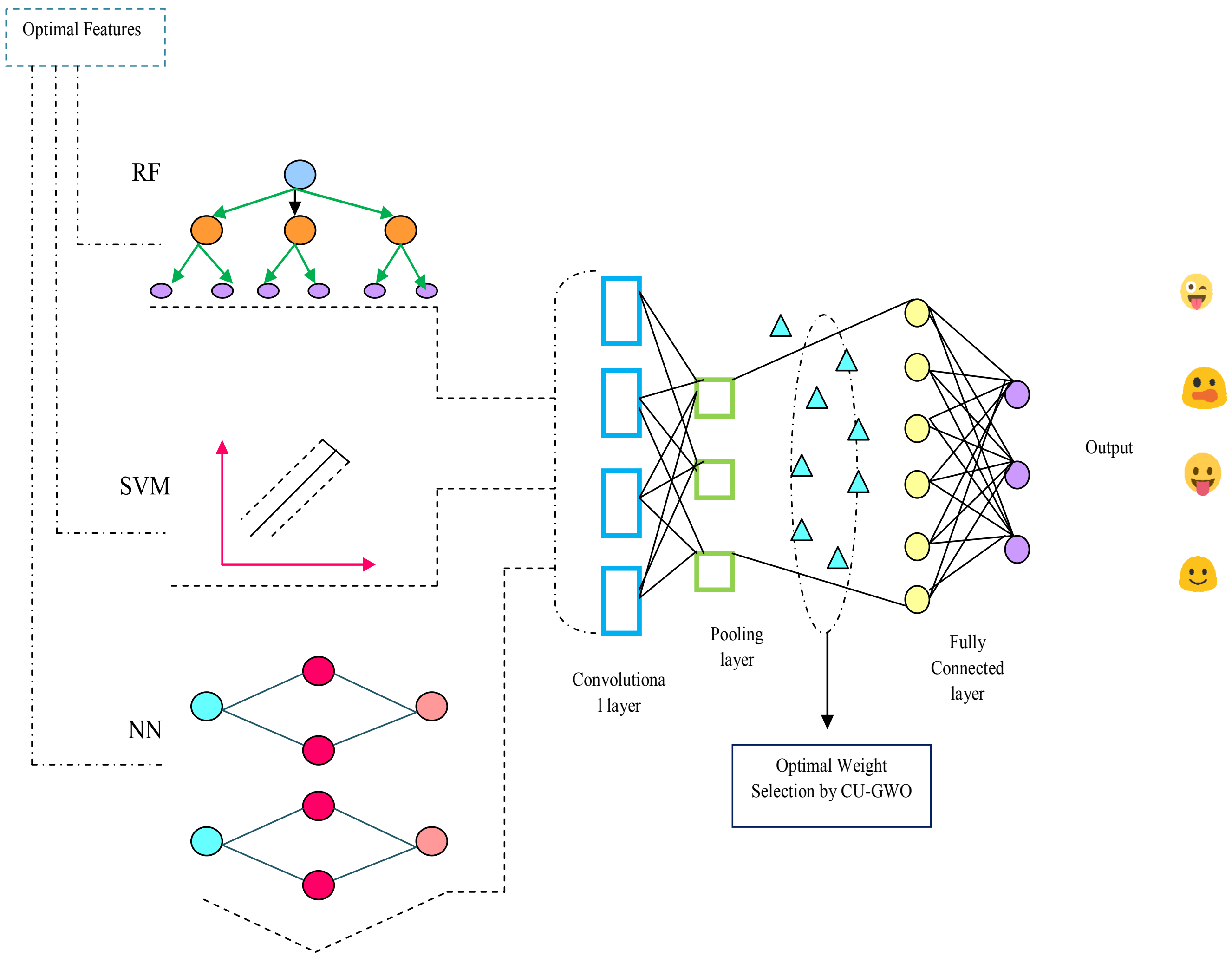
$$A = VT^T \Gamma + hp = 0 \tag{6}$$

The classified output of the SVM is denoted by lu_2.

NN: NN (Mohan et al., 2016; Thorat, 2021) considers the optimal features Fe as input, which is represented in equation (7).

$$Fe = \{Fe_1, Fe_2, ..., Fe_{mu}\} \tag{7}$$

Figure 3 Proposed ensemble technique (see online version for colours)



It (Sarhrouni et al., 2012) includes ‘output, hidden and input layers. Here, $hl^{(\Theta)}$ is defined by equation (8). The term $\hat{Q}_{\hat{o}}$ is defined by equation (9). Further, the error between actual and predicted values is evaluated as in equation (10), which has to be lessened.

$$hl^{(\Theta)} = AF\left(We^{(\Theta)}_{(Bi)} + \sum_{j=1}^{nj} We^{(\Theta)}_{(ji)} Fe \right) \tag{8}$$

$$\hat{Q}_{\hat{o}} = AF\left(We^{(Q)}_{\left(B\hat{O}\right)} + \sum_{j=1}^{nj} We^{(\Theta)}_{\left(i\hat{O}\right)} hl^{(\Theta)} \right) \tag{9}$$

$$Er^{*} = \left\{ \underset{We^{(\Theta)}_{(Bi)},\, We^{(\Theta)}_{(ji)},\, We^{(Q)}_{\left(B\hat{O}\right)},\, We^{(Q)}_{\left(i\hat{O}\right)}}{\arg\min} \right\} = \sum_{1}^{n_G} \left| Q_{\hat{O}} - \hat{Q}_{\hat{O}} \right| \tag{10}$$

Here, ‘mu → total count of features, $hl^{(\Theta)}$ → hidden layer output, AF → activation function, $\hat{i}$ and j → neurons of hidden and input layers, $We^{(\Theta)}_{(Bi)}$ → bias weight to $\hat{i}$ – th hidden neuron, $n_{\hat{i}}$ → count of input neurons, $We^{(\Theta)}_{(ji)}$ → weight from *j*-th input neuron to $\hat{i}$ – th hidden neuron, $\hat{Q}_{\hat{o}}$ → network output, $\hat{O}$ → output neurons, n_h → hidden neuron count, $We^{(Q)}_{\left(B\hat{O}\right)}$ → output bias weight to $\hat{O}$ – th output layer, $We^{(\Theta)}_{\left(i\hat{O}\right)}$ → weight from $\hat{i}$ – th hidden layer to $\hat{O}$ – th output layer, n_G → count of output neuron, $\hat{Q}_{\hat{o}}$ and $Q_{\hat{o}}$ → predicted and actual output’ respectively.

The classified output of the NN model is denoted by lu_3.

The outputs of these classifiers $\left(lu_1 + lu_2 + lu_3 = L\right)$ are then given as input to DCNN for final classification.

Optimised DCNN: DCNN (Gu et al., 2018; Rao et al., 2019; Darekar and Dhande, 2019) include three varied layers such as, ‘convolutional layer, pooling layer and fully-connected layers’. All neurons are connected with adjacent neurons in prior layer. At position (r,t) in *l*-th layer of related *w*-th feature map, the features are computed as per equation (11). The activation value related to $B^{l}_{r,t,w}$ is evaluated as given in equation (12).

$$B^{l}_{r,t,w} = W^{l^T}_{w} PI^{l}_{r,t} + D^{l}_{w} \tag{11}$$

$$act^{l}_{r,t,w} = act\left(B^{l}_{r,t,w}\right) \tag{12}$$

Pooling layer carries out down sampling functions attained from convolutional layers. For each $pool(\bullet)$ related to $act^{l}_{m,h,w}$, the $C^{l}_{r,t,w}$ value is computed as specified in equation (13).

$$C^{l}_{r,t,w} = pool\left(act^{l}_{m,h,w}\right), \forall (m,h) \in NN_{r,t} \tag{13}$$

In the output layer of CNN, the prediction results occur. The term $Loss$ is computed shown as in equation (14).

$$Loss = \frac{1}{mu} \sum_{h=1}^{nu} l\left(\theta; C^{(h)}, F^{(h)}\right) \tag{14}$$

The general constraint related with w^{l}_{w} and D^{l}_{w} is specified by θ. Here, there exist nu counts of output–input relation $\left\{\left(PI^{(h)}, C^{(h)}\right); h \in [1, \dots, nu]\right\}$.

The parameters in DCNN are described below.

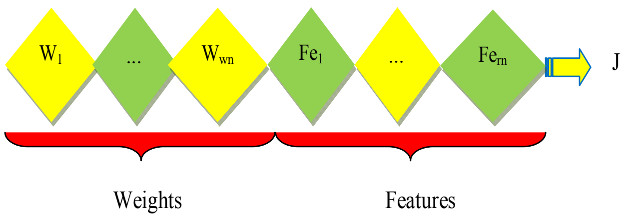
w^{l}_{w} → weight that is optimised via CU-GWO model, D^{l}_{w} → bias term of *w*-th filter related to *l*-th layer, $PI^{l}_{r,t}$ → patch input at centred location (r,t) of *l*-th layer, $act^{l}_{m,h,w}$ → activation value, $B^{l}_{r,t,w}$ → convolutional features, $pool(\bullet)$ → pooling function, $NN_{r,t}$ → neighbourhood near location (r,t) $Loss$ → CNN loss, $PI^{(h)}$ → *h*-th input feature, $C^{(h)}$ → labels and $F^{(h)}$ → output.

The DCNN gives the final detection output indicating the presence of sarcasm.

3.3.2 *Proposed CU-GWO algorithm for optimisation process*

In this work, the DCNN weights denoted by (*W*) as well as the derived features denoted by (*Fe*) are optimally chosen via CU-GWO scheme. Figure 4 shows the representation of solution wherein, wn represents the total count of DCNN weights and rn represent the total count of features.

Figure 4 Solution encoding (see online version for colours)



The existing GWO (Mirjalili et al., 2014; Vinolin, 2019; Roy, R.G. and Ghoshal, 2020) involves a variety of advantages; nevertheless, it endures from certain drawbacks. Therefore, the concept of EHO (Wang et al., 2015) is mingled with it to introduce a new algorithm termed as CU-GWO. Hybridised optimisation schemes are said to be capable of solving different search issues (Beno et al., 2014; Thomas and Rangachar, 2018; Devagnanam and Elango. 2020; Shareef and Rao, 2018, Roy, 2020, 2019). The steps followed in the proposed CU-GWO are as follows.

The wolves α, β and γ are the most important wolves, which concern on hunting process. Amongst these, α is the leader that takes decisions on hunting procedure, sleeping place, time for awakening, etc., while, β and γ takes the 2nd and 3rd level and they aid α in decision making. Accordingly, the last set of wolves is δ that focuses on eating. The encircling features of wolves are formulated as in equation (16). Conventionally, the encircling position is updated based upon position vector of wolves, however, as per developed CU-GWO model, the update is performed based on the clan update of EHO model as in equation (16), where ϕ lies among [0, 1] and it signify the term for computing the impact of clan matriarch on elephants new position, '$\delta \in [0,1]$ signify the likelihood of elephants to move towards clan centre, $\Psi \in [0,1]$ signify the likelihood of elephants to wander (walk in random direction), $rand$ signify random integer, C_{ij}^{it} signify clan centre and m_i^{it} signify matriarch'.

$$Z = \left|B, J_p(it) - J(it)\right| \tag{15}$$

$$J_{ij}^{it+1} = J_{ij}^{it} + \phi \cdot \left(m_{ij}^{it} - J_{ij}^{it}\right) + \delta * \left(C_{ij}^{it} - J_{ij}^{it}\right) \tag{16}$$

The numerical method for depicting the hunting quality of wolf is shown in equation (19) to equation (24). Accordingly, R signify coefficient vectors and it is computed as per equation (17), J_p point out position vector of prey, J point out position vector of wolves and it point out present iteration and Z refers to the distance, B denotes the random vector that lies among 0 and 1 and it is shown in equation (18). In equation (17), constraint $\hat{b}$ lies between 2 to 0 and ra_1 and ra_2 points out the random vectors amongst [0, 1] and $it_{\max}$ point out the maximal iteration.

$$R = 2\hat{b} \cdot ra_1 - \hat{b} \tag{17}$$

$$B = 2ra_2 \tag{18}$$

As per developed model, the final update takes place by taking the harmonic mean as revealed in equation (25). The pseudocode of CU-GWO approach is given in Algorithm 1.

$$Z_\alpha = \left|B_1 J_\alpha - J\right| \tag{19}$$

$$Z_\beta = \left|B_2 J_\beta - J\right| \tag{20}$$

$$Z_\lambda = \left|B_3 J_\gamma - J\right| \tag{21}$$

$$J_1 = \left|J_\alpha - R_1 \cdot \left(Z_\alpha\right)\right| \tag{22}$$

$$J_2 = \left|J_\beta - R_2 \cdot \left(Z_\beta\right)\right| \tag{23}$$

$$J_3 = \left|J_\gamma - R_3 \cdot \left(Z_\gamma\right)\right| \tag{24}$$

$$J(it+1) = \frac{3}{\left[\sum \frac{1}{J_1 + J_2 + J_3}\right]} \tag{25}$$

Algorithm 1: Developed EC + CU-GWO model

Initialisation
Compute the fitness as in equation (1)
Assign J_α as best search agents
Assign J_β as 2nd best search agents
Assign J_δ as 3rd best search agents
While $(it < it_{\max})$
  For every wolf
  Compute the encircling behaviour based on EHO clan update as shown in equation (16)
  Update position based on proposed GWO formulation as per equation (25)
  End for
  Update $\hat{b}$, R and B
  Compute the fitness as in equation (1)
  Update J_α J_β and J_δ
  $it = it + 1$
End while
Return J_α

4 Results and discussions

4.1 Simulation setup

This work uses **Python** for evaluating the enhancement of the EC + CU-GWO based sarcasm detection model. Accordingly, the performance of the adopted approach was measured over extant models such as EC + WOA (Mirjalili and Lewis, 2016), EC + GWO (Mirjalili et al., 2014), EC + EHO (Wang et al., 2015), NN (Ren et al., 2018) and SVM (Bouazizi and Ohtsuki, 2016) with respect to 'accuracy, sensitivity, specificity, precision, F-measure, MCC, FNR, FPR, FDR, NPV and Rand Index'. Here, the performance analysis was performed by altering the testing LR that ranges from 60, 70 and 80. Moreover, convergence analysis was carried out for varied iterations that range from 0, 10, 20, 30, 40 and 50. During analysis, the training and testing rates were set as 70% and 30%, respectively. The experimental parameters used for this evaluation are shown in Table 3.

Table 3 Experimental parameters

Parameters	*Values*
Maximum iteration	25
Population size	5
EHO	$\alpha = 0.5$; $\beta = 0.5$; n_clans=5
WOA	p=0.5; b=1

4.2 Data set details

The data set utilised for this experimentation was downloaded from Kaggle (n.d.). This data set collects data from two news websites: The Onion, which publishes sarcastic versions of current events, and Huff Post, which publishes real news. The data set contains approximately 28,000 headlines, 13,000 of which are sarcastic. Every record in the data set has three attributes:

- is_sarcastic: 1 if the record is sarcastic, otherwise 0.
- headline: the news article's headline.
- article_link: a link to the original news article.

The general statistics of the data set are shown in Table 4.

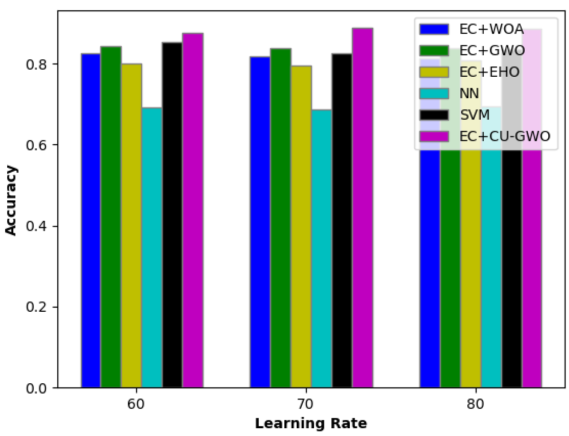
Table 4 General statistics of the data set

Statistic/Data set	*Headlines*	*Semeval*
Records	28,619	3000
Sarcastic records	13,635	2396
Non-sarcastic records	14,984	604
% of pre-trained word embeddings not available	23.35	35.53

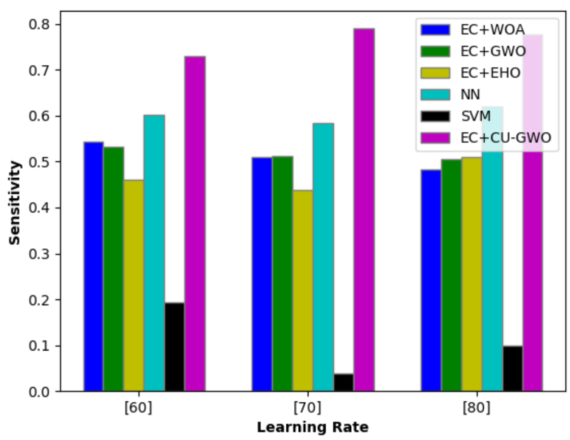
4.3 Performance analysis

The performances of the developed EC + CU-GWO model for sarcasm detection are evaluated over extant models regarding 'positive measures like accuracy, sensitivity, specificity and precision and negative measures like FNR, FPR, FDR and neutral measures like NPV, F-measure, Rand Index and MCC'. The analysis of suggested EC + CU-GWO model is evaluated over EC + WOA, EC + GWO, EC + EHO, NN and SVM for varied LR that ranges from 60, 70 and 80. Consequently, analysis was held and the respective resultants are plotted from Figures 5 to 7. On analysing the entire graphs, the presented EC + CU-GWO model has obtained better outcomes than the compared schemes. This performance efficiency is due to the incorporation of hybrid optimisation in proposed ensemble combination. In fact, the hybrid optimisation ensures a better convergence rate to the defined objectives of a specific problem. As per this work, high-positive values and minimal negative values promise the superior sarcasm detection rate of the model. The performance shown has proven the impact of optimisation in both the feature selection and training of DCNN. Since, optimal training with optimal features ensures the accurate detection of sarcasm existing in the input. On examining Figure 5(c) the specificity of developed approach at 60 th LR is 93, whereas, at 80th LR, the developed model has attained a higher specificity of 95. Likewise, for certain positive measures, the outputs increase with an increase in LR and for certain measures such as (rand index, NPV) the outputs for the proposed model remains same for all LR. In addition, the outputs attained for the negative measure (FNR) at 60th LR is found to be much minimal than the outcomes attained at 90th LR. Especially, excellent outputs for FNR are attained at 70th LR and 80th LR for proposed method when compared to the resultant at 60th LR. Predominantly, from Figure 7(b); finest MCC values are accomplished by adopted approach at 70th LR than the results attained during other LRs. At 70th LR, the specificity of the developed approach is 5.56%, 3.16%, 7.37%, 21.05% and 5.56% better than the values attained by EC + WOA, EC + GWO, EC + EHO, NN and SVM, respectively. Similarly, the negative measures of the adopted model expose minimal values than other models. Thus, the analysis established the enhanced efficacy of the EC + CU-GWO method with better detection outputs.

Figure 5 Performance of proposed approach over compared approaches (a) accuracy (b) sensitivity (c) specificity (d) precision (see online version for colours)

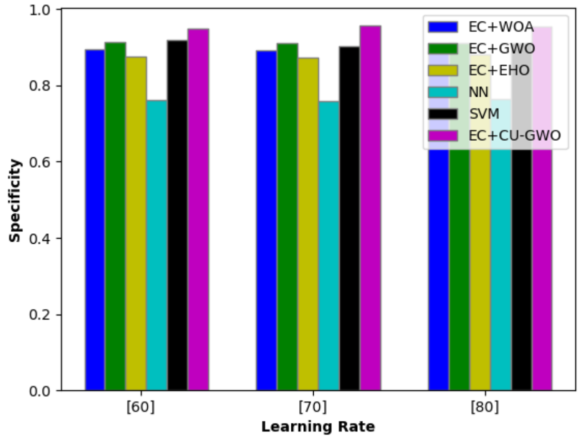


(A)

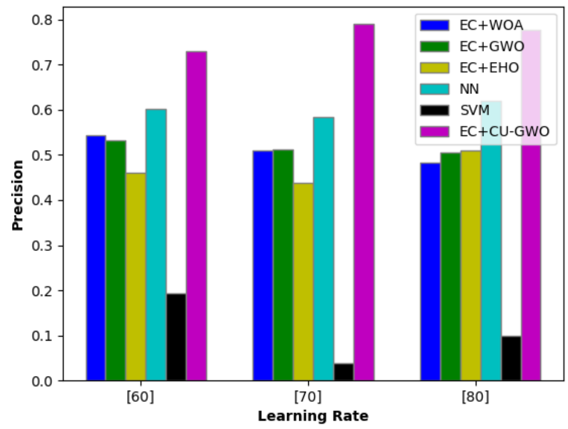


(B)

Figure 5 Performance of proposed approach over compared approaches (a) accuracy (b) sensitivity (c) specificity (d) precision (see online version for colours) (continued)

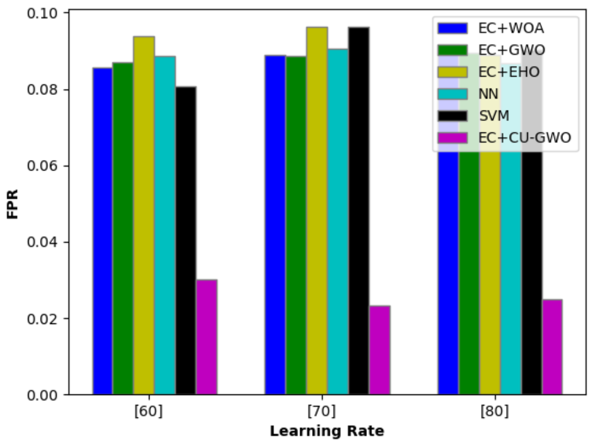


(C)

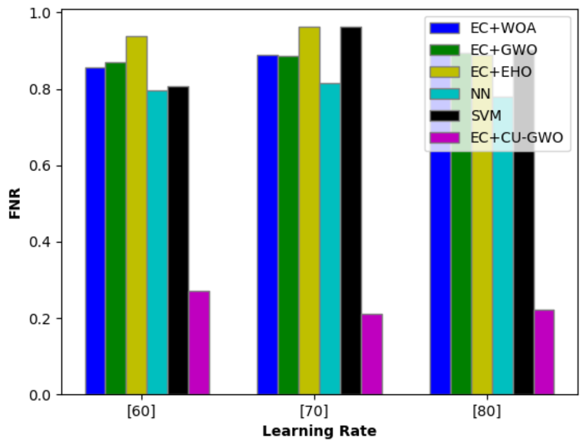


(D)

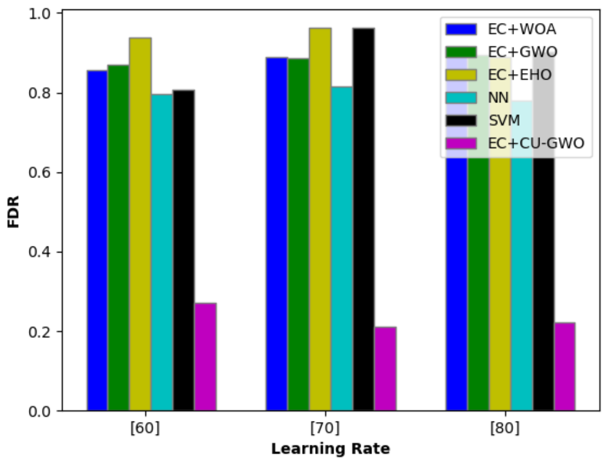
Figure 6 Performance of proposed approach over compared approaches (a) FPR (b) FNR and (c) FDR (see online version for colours)



(A)

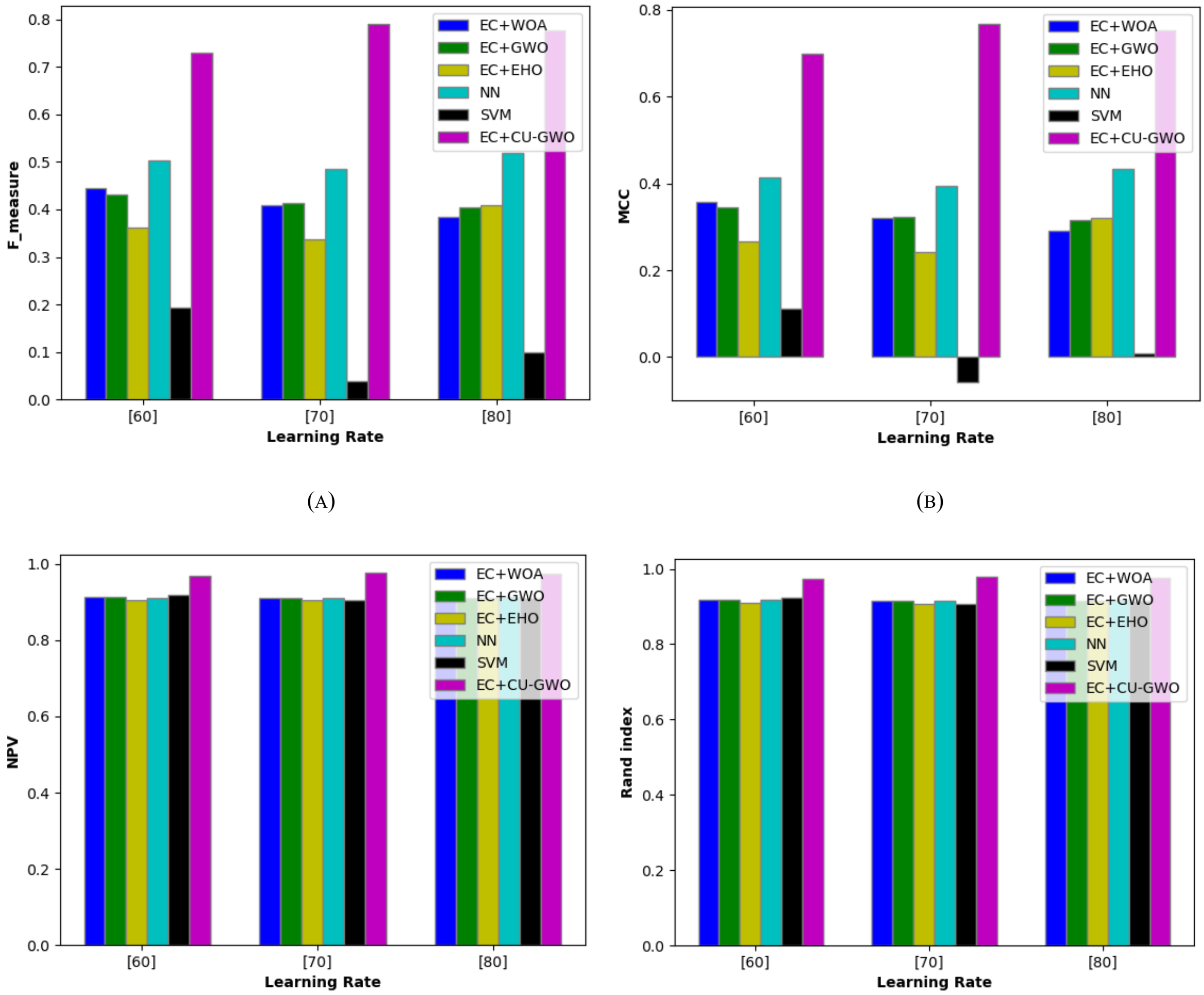


(B)



(C)

Figure 7 Performance of developed approach over compared approaches (a) F measure (b) MCC (c) NPV and (d) Rand index (see online version for colours)



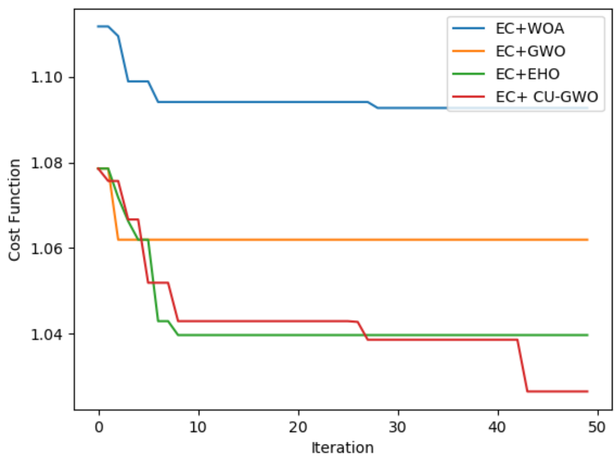
(A) (B) (C) (D)

4.4 *Convergence analysis*

Figure 8 describes the convergence analysis of the adopted model over conventional approaches regarding cost. Here, analysis is performed for a varied number of iterations that ranges from 0, 10, 20, 30, 40 and 50. On noticing the analysis resultants, the developed EC + CU-GWO model has obtained negligible cost for all iterations over conventional approaches. The cost values for the suggested model as well as compared models are found to be lessening with raise in the count of iteration. However, the adopted model has accomplished reduced cost values over the other evaluated schemes. At the initial iterations (from 0 to 42), the cost of the developed model has accomplished a comparatively higher value, while at further iterations; the developed model has converged to a minimal cost value. Especially, at iteration 5, the adopted model is only 3.8% enhanced than the extant WOA model, while at iteration 50, the adopted model is 6.54% enhanced than WOA model. Predominantly, on noticing cost values from Figure 8, the adopted scheme has attained a reduced cost value (1.025) from iteration 42 to 50.

Thus, the overall assessment shows the impact of the developed EC + CU-GWO model; on better convergence results with an increase in iterations.

Figure 8 Convergence analysis of developed scheme over traditional models (see online version for colours)



4.5 Statistical analysis

The statistical analysis of the implemented EC + CU-GWO model over other traditional models for varied scenarios is shown in Table 5. 'As the meta-heuristic schemes are stochastic in nature, every algorithm is executed for number of times to attain the statistic of objective function'. The adopted EC + CU-GWO model demonstrates the superior outcomes when evaluated over conventional schemes such as EC + WOA, EC + GWO and EC + EHO models. From Table 5, the proposed EC + CU-GWO model under median case scenario attains least cost when compared to other models. This proves the detection efficiency of proposed work than the conventional models. Though in some cases, the performance of proposed algorithm seems to be worst; however, the overall performance proves its betterment over other conventional models.

Table 5 Statistical analysis of the adopted model over existing models

Methods	*Standard deviation*	*Mean*	*Median*	*Maximum*	*Minimum*
EC + WOA	0.010576	1.080772	1.077089	1.092697	1.07253
EC + GWO	0.020675	1.036765	1.039681	1.055827	1.014787
EC + EHO	0.020675	1.036765	1.039681	1.055827	1.014787
EC + CU-GWO	0.012903	1.032826	1.026518	1.047669	1.02429

5 Conclusion

This paper has developed a novel sarcasm detection model with four phases. At first, pre-processing was performed and then features like 'information gain, chi-square, mutual information and symmetrical uncertainty-based features' were extracted. Then, optimal features were chosen that were subjected to an ensemble classifier via NN, RF, SVM and DCNN. Especially, the weights of DCNN were optimally tuned via a new CU-GWO algorithm. Thus, the sarcasm was detected. At last, the superiority of the offered scheme was established over the conventional schemes regarding diverse measures. For certain positive measures, the outputs increase with an increase in LR and for certain measures such as (rand index, NPV) the outputs for proposed model remain same for all LR. In addition, the outputs attained for negative measure (FNR) at 60th LR was found to be much minimal than the outcomes attained at 90th LR. Especially, excellent outputs for FNR were attained at 70th LR and 80th LR for proposed method when compared to the resultant at 60th LR. Predominantly, finest MCC values were accomplished by adopted approach at 70th LR than the results attained during other LRs. At 70th LR, the specificity of developed approach was 5.56%, 3.16%, 7.37%, 21.05% and 5.56% better than the values attained by EC + WOA, EC + GWO, EC + EHO, NN and SVM, respectively. Similarly, the negative measures of the adopted model exposed minimal values than other models. Therefore, the supremacy of the introduced approach has been confirmed effectively.

References

Ashwitha, A., Shruthi, G. and Manjunath, T.C. (2020) 'Sarcasm detection in natural language processing', *Materials Today: Proceedings*, Vol. 37, pp.3324–3331.

Beno, M.M., Valarmathi, I.R., Swamy, S.M. and Rajakumar, B.R. (2014) 'Threshold prediction for segmenting tumour from brain MRI scans', *International Journal of Imaging Systems and Technology*, Vol. 24, No. 2, pp.129–137. Doi: 10.1002/ima.22087.

Bharti, S.K., Vachha, B., Pradhan, R.K., Babu, K.S. and Jena, S.K. (2016) 'Sarcastic sentiment detection in tweets streamed in real time: a big data approach', *Digital Communications and Networks*, Vol. 2, No. 3, pp.108–121.

Bouazizi, M. and Ohtsuki, T.O. (2016) 'A pattern-based approach for sarcasm detection on Twitter', *IEEE Access*, Vol. 4, pp.5477–5488.

Cai, X., Zhang, J., Ning, Z., Cui, Z. and Chen, J. (2021) 'A many-objective multistage optimization-based fuzzy decision-making model for coal production prediction', *IEEE Transactions on Fuzzy Systems*, Vol. 29, No. 12 pp.3665–3675.

Creusere, M.A. (1999) 'Theories of adults' understanding and use of irony and sarcasm: applications to and evidence from research with children', *Developmental Review*, Vol. 19, No. 2, pp.213–262.

Cui, Z., Zhang, J., Wang, Y., Cao, Y., Cai, X., Zhang, W. and Chen, J. (2019) 'A pigeon-inspired optimization algorithm for many-objective optimization problems', *Science China Information Science*, Vol. 62, No. 7.

Darekar, R.V. and Dhande, A.P. (2019) 'Emotion recognition from speech signals using DCNN with hybrid GA-GWO algorithm', *Multimedia Research*, Vol. 2, No. 4, pp.12–22.

Del Pilar Salas-Zárate, M., Paredes-Valverde, M.A., Rodriguez-García, M.A., Valencia-García, R. and Alor-Hernández, G. (2017) 'Automatic detection of satire in Twitter: a psycholinguistic-based approach', *Knowledge-Based Systems*, Vol. 128, pp.20–33.

Deliens, G., Antoniou, K., Clin, E. and Kissine, M. (2017) 'Perspective-taking and frugal strategies: evidence from sarcasm detection', *Journal of Pragmatics*, Vol. 119, pp.33–45.

Devagnanam, J. and Elango, N.M. (2020) 'Optimal resource allocation of cluster using hybrid grey wolf and cuckoo search algorithm in cloud computing', *Journal of Networking and Communication Systems*, Vol. 3, No. 1, pp.31–40.

Ebrahimi, M., Yazdavar, A.H. and Sheth, A. (2017) 'Challenges of sentiment analysis for dynamic events', *IEEE Intelligent Systems*, Vol. 32, No. 5, pp.70–75.

Entezari-Maleki, R., Rezaei, A. and Minaei-Bidgoli, B. (2009) 'Comparison of classification methods based on the type of attributes and sample size', *Journal of Convergence Information Technology*, Vol. 4, No. 3, pp.94–102.

Gazeloglu, C. (2020) Prediction of heart disease by classifying with feature selection and machine learning methods', *Progress in Nutrition*, Vol. 22, No. 2, pp.660–667. Available online at: https://www.mattioli1885journals.com/index.php/progressinnutrition/article/view/9830

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T. (2018) 'Recent advances in convolutional neural networks', *Pattern Recognition*, Vol. 77, pp354–377.

Jain, D., Kumar, A. and Garg, G. (2020) 'Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN', *Applied Soft Computing*, Vol. 91.

Justo, R., Corcoran, T., Lukin, S.M., Walker, M. and Torres, M.I. (2014) 'Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web', *Knowledge-Based Systems*, Vol. 69, pp.124–133.

Kaggle (n.d.) *News headlines dataset for sarcasm detection*. Available online at: https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection

Kumar, H.M.K. and Harish, B.S. (2018) 'Sarcasm classification: a novel approach by using content based feature selection method', *Procedia Computer Science*, Vol. 143, pp.378–386.

Kunneman, F., Liebrecht, C. and Van den Bosch, A. (2015) 'Signaling sarcasm: from hyperbole to hashtag', *Information Processing and Management*, Vol. 51, No. 4, pp.500–509.

Lagerwerf, L. (2007) 'Irony and sarcasm in advertisements: effects of relevant inappropriateness', *Journal of Pragmatics*, Vol. 39, No. 10, pp.1702–1721.

Ludlow, A.K., Chadwick, E., Morey, A., Edwards, R. and Gutierrez, R. (2017) 'An exploration of sarcasm detection in children with attention deficit hyperactivity disorder', *Journal of Communication Disorders*, Vol. 70, pp.25–34.

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A.T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H.C. and Sadato, N. (2016) 'The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence', *Neuropsychologia*, Vol. 87, pp.74–84.

Mirjalili, S. and Lewis, A. (2016) 'The whale optimization algorithm', *Advances in Engineering Software*, Vol. 95, pp.51–67.

Mirjalili, S., Mirjalili, S.M. and Lewis, A. (2014) 'Grey wolf optimizer', *Advances in Engineering Software*, Vol. 69, pp.46–61.

Mohan, Y., Chee, S.S., Xin, D.K.P. and Foong, L.P. (2016) 'Artificial neural network for classification of depressive and normal in EEG', *Proceedings of the IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, Malaysia.

Mukherjee, S. and Bala, P.K. (2017) 'Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering', *Technology in Society*, Vol. 48, pp.19–27.

Persicke, A., Tarbox, J., Ranick, J. and Clair, M.S. (2013) 'Teaching children with autism to detect and respond to sarcasm', *Research in Autism Spectrum Disorders*, Vol. 7, No. 1, pp.193–198.

Rao, T.C.S., Ram, S.S.T. and Subrahmanyam, J.B.V. (2019) 'Enhanced deep convolutional neural network for fault signal recognition in the power distribution system', *Journal of Computational Mechanics, Power System and Control*, Vol. 2, No.3, pp.39–46.

Ren, L., Xu, B. and Yang, L. (2020) 'Sarcasm detection with sentiment semantics enhanced multi-level memory network', *Neurocomputing*, Vol. 401, pp.320–326.

Ren, Y., Ji, D. and Ren, H. (2018) 'Context-augmented convolutional neural networks for twitter sarcasm detection', *Neurocomputing*, Vol. 308, pp.1–7.

Roy, R.G. (2019) 'Rescheduling based congestion management method using hybrid grey wolf optimization-grasshopper optimization algorithm in power system', *Journal of Computational Mechanics, Power System and Control (JCMPS)*, Vol. 2, No. 1, pp.9–18.

Roy, R.G. (2020) 'Economic dispatch problem in power system using hybrid PSO and enhanced bat optimization algorithm', *Journal of Computational Mechanics, Power System and Control (JCMPS)*, Vol. 3, No. 3, pp.27–33.

Roy, R.G. and Ghoshal, D. (2020) 'Grey wolf optimization-based second order sliding mode control for inchworm robot', *Robotica*, Vol. 38, No. 9, pp.1539–1557.

Sanders, R.E. (2013) 'The duality of speaker meaning: what makes self-repair, insincerity, and sarcasm possible', *Journal of Pragmatics*, Vol. 48, No. 1, pp.112–122.

Sarhrouni, E., Hammouch, A. and Aboutajdine, D. (2012). Application of symmetric uncertainty and mutual information to dimensionality reduction and classification of hyperspectral images', *International Journal of Engineering and Technology*, Vol. 4, pp.268–276.

Schouten, K. and Frasincar, F. (2016) 'Survey on aspect-level sentiment analysis', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 3, pp.813-830.

Shareef, S.K.M. and Rao, R.S. (2018) 'A hybrid learning algorithm for optimal reactive power dispatch under unbalanced conditions', *Journal of Computational Mechanics, Power System and Control*, Vol. 1, No. 1, pp.26–33.

Son, L.H., Kumar, A., Sangwan, S.R., Arora, A., Nayyar, A. and Abdel-Basset, M. (2019) 'Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network', *IEEE Access*, Vol. 7, pp.23319–23328. Doi: 10.1109/ACCESS.2019.2899260.

Sonawane, S.S. and Kolhe, S.R. (2020) 'TCSD: term co-occurrence based sarcasm detection from Twitter trends', *Procedia Computer Science*, Vol. 167, pp.830–839.

Sulis, E., Farías, D.H., Rosso, P., Patti, V. and Ruffo, G. (2016) 'Figurative messages and affect in Twitter: differences between #irony, #sarcasm and #not', *Knowledge-Based Systems*, Vol. 108, pp.132–143.

Thomas, R. and Rangachar, M.J.S. (2018) 'Hybrid optimization based DBN for face recognition using low-resolution images', *Multimedia Research*, Vol. 1, No. 1, pp.33–43.

Thomas, R. and Rangachar, M.J.S. (2019) 'Fractional rider and multi-kernel-based spherical SVM for low resolution face recognition', *Multimedia Research*, Vol. 2, No. 2, pp.35–43.

Thorat, V.A. (2021) 'Cloud intrusion detection using modified crow search optimized based neural network', *Journal of Networking and Communication Systems*, Vol. 4, No. 2.

Uchiyama, H., Seki, A., Kageyama, H., Saito, D.N., Koeda, T., Ohno, K. and Sadato, N. (2006) 'Neural substrates of sarcasm: a functional magnetic-resonance imaging study', *Brain Research*, Vol. 1124, No. 1, pp.100–110.

Vinolin, V. (2019) 'Breast cancer detection by optimal classification using GWO algorithm', *Multimedia Research*, Vol. 2, No. 2, pp.10–18.

Wang, G.G., Deb, S. and Coelho, L. (2015) 'Elephant herding optimization', *Proceedings of the 3rd International Symposium on Computational and Business Intelligence (ISCBI)*, IEEE, Bali, Indonesia. Doi:10.1109/ISCBI.2015.8.

Zhao, J., Chen, D., Xiao, R., Cui, Z., Wang, H. and Lee, I. (2022) 'Multi-strategy ensemble firefly algorithm with equilibrium of convergence and diversity', *Applied Soft Computing,* Vol. 123. Doi: 10.1016/j.asoc.2022.108938.

Nomenclature

Abbreviation	*Description*
Bi-LSTM	Bidirectional Long Short-Term Memory
CU-GWO	Clan Updated Grey Wolf Optimisation
CNN	Convolutional Neural Network
convNet	Convolution Neural Network
DNN	Deep Neural Network
DT	Decision Tree
DCNN	Deep Convolutional Neural Network
EHO	Elephant Herding Optimisation
EC	Ensemble Classification
FDR	False Discovery Rate
FPR	False Positive Rate
FNR	False Negative Rate
GWO	Grey Wolf Optimisation
GLoVe	Global Vectors For Word Representation
LR	Learning Rate
MCC	Matthews Correlation Coefficient
MI	Mutual Information
NB	Naïve Bayes
NPV	Negative Predictive Value
NLP	Natural Language Processing
NN	Neural Network
RF	Random Forest
SVM	Support Vector Machine
sAtt-BLSTM	Soft Attention-Based Bidirectional LSTM
WOA	Whale Optimisation