Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.027502
Article

Tech Science Press

Optimal Weighted Extreme Learning Machine for Cybersecurity Fake News Classification

Ashit Kumar Dutta[1,*], Basit Qureshi[2], Yasser Albagory[3], Majed Alsanea[4], Manal Al Faraj[1] and Abdul Rahaman Wahab Sait[5]

[1]Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Ad Diriyah, Riyadh, 13713, Kingdom of Saudi Arabia
[2]Department of Computer Science, Prince Sultan University, Riyadh, 11586, Kingdom of Saudi Arabia
[3]Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif, 21944, Kingdom of Saudi Arabia
[4]Department of Computing, Arabeast Colleges, Riyadh, 11583, Kingdom of Saudi Arabia
[5]Department of Archives and Communication, King Faisal University, Al Ahsa, Hofuf, 31982, Kingdom of Saudi Arabia
*Corresponding Author: Ashit Kumar Dutta. Email: adotta@mcst.edu.sa
Received: 19 January 2022; Accepted: 23 March 2022

Abstract: Fake news and its significance carried the significance of affecting diverse aspects of diverse entities, ranging from a city lifestyle to a country global relativity, various methods are available to collect and determine fake news. The recently developed machine learning (ML) models can be employed for the detection and classification of fake news. This study designs a novel Chaotic Ant Swarm with Weighted Extreme Learning Machine (CAS-WELM) for Cybersecurity Fake News Detection and Classification. The goal of the CAS-WELM technique is to discriminate news into fake and real. The CAS-WELM technique initially pre-processes the input data and Glove technique is used for word embedding process. Then, N-gram based feature extraction technique is derived to generate feature vectors. Lastly, WELM model is applied for the detection and classification of fake news, in which the weight value of the WELM model can be optimally adjusted by the use of CAS algorithm. The performance validation of the CAS-WELM technique is carried out using the benchmark dataset and the results are inspected under several dimensions. The experimental results reported the enhanced outcomes of the CAS-WELM technique over the recent approaches.

Keywords: Cybersecurity; cybercrime; fake news; data classification; machine learning; metaheuristics

1 Introduction

Online data is often accessible as a result of few clicks away. With the unique independence provided to users for sharing stories, the complexity to describe the root of false data increases gradually. The existence of dramatic headlines and clickbait titles is at its highest point that assists in the broadcast of inaccurate and unprofessional news in response to advertising revenues. User, wants to be part of this hot discussion or topic, adapt the innovative message with intention or by mistake that eventually results in the distribution of rumor on the internet. Fake news is inscribed for a hoax that leads to political or gains or financial spreading data disguised as propaganda [1], one might be utilized to influence public perception towards falseness. Even this encourages the beliefs and people ideology to some range that might create several damages [2]. This persuading is popular when a news story breaks out, whereby the supporter usually tends to share data in its complete originality, while the one opinion doesn't bring into line with the information mentioned resorting to share that similar data with few adjustments. Currently, media outlets are the only information resources. Specific contribution in news sharing has significantly developed over the last decade where it become ever more complex to discriminate news that originate from a reliable source from the one that is invented [3]. Consequently, fake news has gained several interests recently by organizations like Google, Twitter, Facebook, and by various authors, who are making continuous attempts in opposing the spread of fake stories. Fig. 1 illustrates the platform to detect fake news.

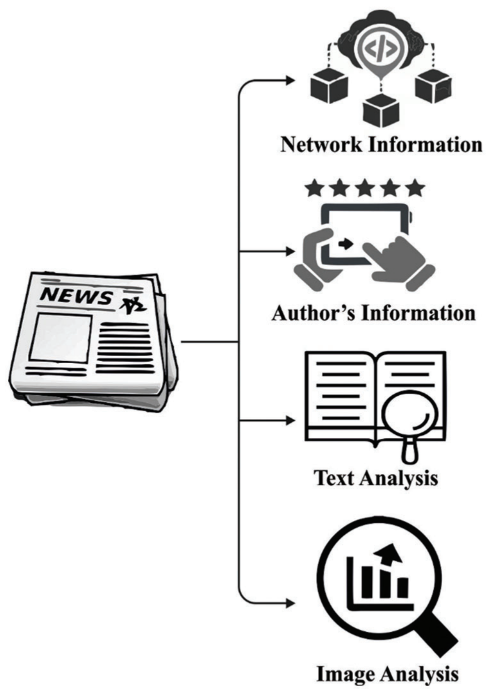

Figure 1: Platform to detect fake news

Artificial intelligence (AI) technique is the evolving technology that has transformed the view at business problems [4]. An increasing amount of businesses are transforming to innovative analysis and machine learning to resolve problems. With this development, natural language processing (NLP) describes great potential for business that is concerned with understanding human sentiment via the current information. NLP functions with each kind of social and natural communication, involving text, audio, and video. In order to identify trends and many valuable patterns in the textual data set, text mining assisted to perform in this way [5]. In present market setting, strategic use of NLP assist business to obtain relative benefits. AI and NLP assist in combating the large unstructured data of various fields involving education, healthcare, business sectors, fake news, trust and security, opinion from the public in the government sector [6]. The NLP assists human-to-machine communication very efficiently that sequentially improves the overall efficiency and decision-making of the businesses. The NLP relates to how individual interacts, that consist of emotions, speech, and text. Fake news detection has gained much consideration in the NLP research field to mitigate the time-consuming human activity and burdensome data verification [7]. Despite that, the process of estimating the validity of news remains a challenge even for automatic systems.

Kumar et al. [8] gather 1356 news samples from different clients through media sources and Twitter including PolitiFact and construct various data sets for the fake and real news stories. We compared many advanced methods including attention mechanism, convolution neural network (CNN), long short term memory (LSTM), and ensemble approaches. Roy et al. [9] developed deep learning (DL) algorithms to identify fake news and classify them to the pre-determined fine-grained classes. Firstly, we designed CNN and bidirectional LSTM (Bi-LSTM) based systems. The representation attained from these two methods is given to a multilayer perceptron (MLP) for the last classification.

Aslam et al. [10] presented an ensemble-based DL method for classifying news as real or fake. Because of the nature of dataset traits, two DL methods have been employed. For the textual attributes "statement," Bi-LSTM-gated recurrent unit (GRU)-dense DL method has been utilized, for the residual characteristics, dense DL algorithm has been employed.

In Agarwal et al. [11], researchers have experimented and discussed word embedding (GloVe) for text pre-processing to establish lingual relationships and create a vector space of words. The presented method is the combination of CNN and recurrent neural network (RNN) frameworks that have accomplished standard outcomes in predicting fake news, with the effectiveness of word embedding complementing the overall method. Furthermore, to guarantee the prediction quality, several model parameters were recorded and tuned for the optimal result.

Khanam et al. [12] make research analytics based fake news detection and examine the conventional machine learning (ML) methods for choosing the best, to construct a method of a product using supervised ML method, which could categorize fake news as false or true, by utilizing python scikit-learn, NLP for text analysis. Bangyal et al. [13] developed a precise model for SA of fake news. The fake news datasets contain fake news; the study initiates by data pre-processing (replaces the stemming, tokenization, noise removal, and missing value). The study employed a semantic method with inverse document frequency and term frequency weighting for representing information. In the evaluation and measuring stage, we employed 8 ML approaches.

This study designs a novel Chaotic Ant Swarm with Weighted Extreme Learning Machine (CAS-WELM) for Cybersecurity Fake News Detection and Classification. The goal of the CAS-WELM system is to discriminate news into fake and real. The CAS-WELM technique initially pre-processes the input data and Glove technique is used for word embedding process. Then, N-gram based feature extraction technique is derived to generate feature vectors. Lastly, WELM model is applied for the detection and classification of fake news, in which the weight value of the WELM model can be optimally adjusted by the use of CAS algorithm. The performance validation of the CAS-WELM technique is carried out using the benchmark dataset and the results are inspected under several dimensions.

2 The Proposed Model

In this study, a novel CAS-WELM technique has been developed for Cybersecurity Fake News Detection and Classification. The CAS-WELM technique mainly intends to discriminate news into fake and real. The CAS-WELM technique undergoes different stages of operations namely pre-processing, Glove based word embedding, N-gram based feature extraction, WELM based classification, and CAS based parameter optimization. Besides, the weight value of the WELM model can be optimally adjusted by the use of CAS algorithm.

2.1 Pre-processing

The data set is considered into two groups, true category, and false category. Data visualization assists in comprehending comparative data mean by demonstrating information in visual contexts, namely graphs or maps. This makes it easy to spot outliers, trends, and patterns in massive datasets by creating the data to analyze for the human mind. The data set is categorized into two classes, original and fake news. The fake news class is denoted as '0' and true news class is denoted as '1'. When certain words exist in the group of a *corpus*, then the word is removed [14]. Data pre-processing is a major phase that includes data manipulation beforehand it is implemented, to improve efficacy. It includes data transformation and cleansing. To remove the stop word from the sentence, the text can be separated into words, and then it is verified to understand whether the word exist in the Natural Language Toolkit (NLTK) list of stop words. Stemming represents the extraction of word root or stems form that may or may not completely reflects semantic intellectual. The procedure of lemmatization is the decrease of inflectional format generally useful word-to common form. Glove embedding and Keras embedding layer, utilized to train NN system on textual information. This is a flexible layer, utilized for loading pre-trained GloVe embedding of hundred dimensions.

2.2 N-gram Based Feature Extraction

Consider $x_i \in R^d$ represent the word vector for *ith* word in a sentence of d dimension. Where $x \in R^{L\times d}$ signify the input sentence using length L. Take k as the filter length, also $m \in R^{k\times d}$ represent a filter for the convolutional process. For all the location, j in the sentence, a window vector w_j consist of k successive word vectors are evaluated,

$$w_j = \left[x_j, x_{j+1}, \ldots, x_{j+k-1}\right] \tag{1}$$

Now, the comma represents row vector concatenation. A filter m integrates to the window vector $(k - grams)$ and all the locations in an approach to construct a feature map $c \in R^{L-k+1}$; all the elements c_j of feature maps for window vector w_j is generated by:

$$c_j = f\left(w_j \circ m + b\right) \tag{2}$$

whereas $\circ$ indicates element-by-element multiplication, $b \in R$ show a bias term and f denotes a non-linear conversion with probable kinds such as sigmoid, hyperbolic tangent, linear, rectified linear unit (ReLU), softmax, and so on. In this case, ReLU is employed. A filter amount is utilized for producing feature map [15]. For n filters of equivalent size, the generated n feature map is rearranged as feature representation for all the window w_j,

$$W = [c_1; c_2; \ldots; c_n] \tag{3}$$

Here, c_i denotes the feature map generated using the *ith* filter and Semicolon signifies column vector concatenation. All the rows W_j of $W \in R^{(L-k+1)\times n}$ represent the feature depiction generated from n feature for window vector at location provided by j.

2.3 WELM Based Classification

Beforehand elaborating on the WELM, firstly presented the fundamental extreme learning machine (ELM). Using the mapping datasets $\{\chi_i', y_j\} \in \Re^p \times \Re^c (i = 1, 2, \ldots, n)$, the output of generalized single layer feed forward network (SLFN) using activation function $h(x')$ and q hidden node can be formulated by using the following equation. Fig. 2 demonstrates the structure of WELM.

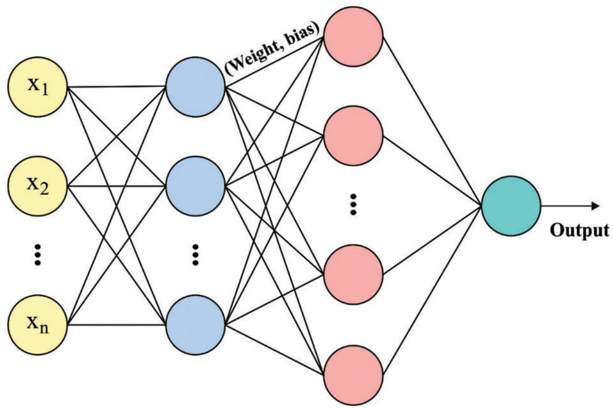

Figure 2: Structure of WELM

$$o_i = \sum\nolimits_{k=1}^{q} \beta_k h_k\left(\chi_i'\right) = \sum\nolimits_{k=1}^{q} \beta_k h\left(w_k, b_k, \mathrm{x}_\mathrm{i}'\right), \tag{4}$$

In which $i = 1, 2, \ldots, n, w_k = [w_{k1}, w_{k2}, \ldots, w_{kp}]^T$ characterizes the input weight connect the kth hidden and input nodes, b_k signifies the bias of *kth* hidden node, $\beta_k = [\beta_{k1}, \beta_{k2}, \ldots, \beta_{kc}]^T$ shows the output weight linking the kth hidden and output nodes, and 0_i represents the predicted output of ith sample. The widely employed activation function in ELM includes multiquadric function, Gaussian RBF function, sigmoid function, and hard limit function [16].

$$H\beta = O, \tag{5}$$

In which H represent the hidden neuron output matrix of SLFN

$$H = H\left(w_1, \ldots, w_q, b_1, \ldots, b_q, x_1', \ldots, x_n'\right) = \begin{bmatrix} h(x_1') \\ \vdots \\ h(x_n') \end{bmatrix} \tag{6}$$

$$= \begin{bmatrix} h\left(w_1, b_1, x_1'\right) & \cdots & h\left(w_q, b_q, x_1'\right) \\ \vdots & \ddots & \vdots \\ h\left(w_1, b_1, x_1'\right) & \cdots & h\left(w_q, b_q, x_1'\right) \end{bmatrix}_{n \times q}$$

Here, the *ith* row of H represent the output of hidden node regarding the input samples x_i', and the kth column of H shows the output of kth hidden node regarding the input sample $x_1', x_2', , .., x_n'$.

β indicates the weight matrix linking the output and hidden layers, as follows

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_q^T \end{bmatrix}_{q \times c} \tag{7}$$

O represents the predicted label matrix, and all the rows represent the output vector of single instance. O is determined by

$$O = \begin{bmatrix} o_1^T \\ \vdots \\ T_{O_n} \end{bmatrix} = \begin{bmatrix} o_{11} & \cdots & 0_{1c} \\ \vdots & \ddots & \vdots \\ o_{n1} & \cdots & 0_{nc} \end{bmatrix} \tag{8}$$

The aim of trained SLFN is to reduce the output errors, that is, approximate the input sample with zero error

$$\sum_{i=1}^{n} \|0_i - y_i\| = \|O - Y\| = 0 \tag{9}$$

Whereas $y = \begin{bmatrix} y_1^T \\ \vdots \\ Ty_n \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1c} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nc} \end{bmatrix}$ represents the target output matrix.

$$H\beta = Y \tag{10}$$

Aimed at ELM, the bias b_k of hidden neurons and the weight w_k of input connection is independently and arbitrarily chosen [17]. When this parameter is allocated, Eq. (10) is transformed to linear method and the β output weight matrix is systematically defined by detecting the least-square solutions of linear method as

$$\min_{\beta} \|H\beta - Y\| \tag{11}$$

The optimum solution of Eq. (11) is

$$\hat{\beta} = H^{\dagger}Y = \left(H^T H\right) \tag{12}$$

Whereas $H\dagger$ means the Moore-Penrose generalized inverse of hidden neuron output matrix H. The attained $\hat{\beta}$ could guarantee minimally trained error, attain optimum generalization capability, and prevent plunging to local optimal because $\hat{\beta}$ is exclusive.

$$f(x') = h(x')\hat{\beta} = h(x')H^{\dagger y} \tag{13}$$

While constructing the ELM classification, we determine a $n \times n$ diagonal matrix W, that diagonal component W_{ii} represents the weight of trained instance χ_i'. Accurately, when χ_i' belonging to the majority class, the W_{ii} weight is comparatively lesser when compared to the samples that belong to the minority class.

$$\hat{\beta} = H^{\dagger}Y = (H^T WH)^{-1} H^T WT \tag{14}$$

Next, Eq. (13) becomes

$$f(x') = h(x')\hat{\beta} = h(x')(H^T WH)^{-1} H^T WT \tag{15}$$

Mainly, it consists of two systems to assign the weight to the sample of the two classes:

$$W1 = W_{ii} = \begin{cases} 1/n_P & \textit{if } x_i' \in \text{minority class} \\ 1/n_N & \textit{if } x_i' \in \text{majority class} \end{cases} \tag{16}$$

Or

$$W2 = W_{ii} = \begin{cases} 0.618/n_P & if\ x_i' \in \text{minority class} \\ 1/n_N & if\ x_i' \in \text{majority class} \end{cases}, \tag{17}$$

Here, Wl and $W2$ denotes weighting systems, n_P & n_N indicates the amount of instances of the minority and majority classes, correspondingly.

2.4 Parameter Optimization Using CAS Algorithm

For tuning the weight values of the WELM model, the CAS is used. Recently, a SI optimization method named CAS approach is presented for solving the optimization issue according to chaos concept [18]. The CAS algorithm is mathematically modelled by the following equation:

$$y_i(t) = y_i(t-1)^{(1+r_i)},$$

$$z_{id}(t) = \Delta \exp((1 - \exp(-ay_i(t)))(3 - \Psi_d \Delta)) -$$

$$\frac{7.5}{\Psi_d \times V_i} + \exp(-2ay_i(t) + b) \times (pbest_d(t-1) - z_{id}(t-1)), \tag{18}$$

In which $y_i(t)$ represent the organization parameter of the CAS and $\Delta = z_{id}(t-1) + 7.5/(\Psi_d \times \varphi_i)$. It handles chaotic behavior of one ant. r_i indicates the organization variances of one ant that is a positive constant lesser than 1. Ψ_i defines the search array of *ith* ant in *dth* dimension. $\varnothing_i$ control the moving proportion of *ith* ant search region. pbest $(t-1)$ denotes the optimal location that the single ant and neighbors have established with $t-1$ time step. Now the neighbor is fixed to be global neighbor; viz., each ant is the neighbor of one another. Usually, The ant exchanges data through direct or indirect transmission models. Owing to the efficient transmission, the effect of organization becomes robust as time changes. At last, each ant walks through the optimal route to forage food. As time grows, the impact of the organization parameter $y_i(t)$ on the behavior of all the ants become strong through the organization variable r_i.

Lastly, with the impact of $pbest_d(t-1)$ and $y_i(t)$, the state of $z_{id}(t)$ would converge to the global optimal location. r_i and Ψ_d represent significant variables. r_i has an impact on the convergence rate of CAS approach. When r_i is smaller, the convergence rate of the CAS approach would be slower and the implementation time would take time. When r_i is larger, the convergence rate of CAS approach would be faster thereby the optimum solution mightn't be established. When r_i is fixed to be 0, the behavior of ant would be chaotic continually and the CAS approach could not converge to a certain location. Moreover, slight variation of organization impact is chosen, r_i is fixed to be $0 \le r_i \le 0.5$. The actual equation of r_i based on the runtime and certain issues.

To support ant to have a distinct organization variable, fix $r_i = 0.1 + 0.2 \times rand$, whereas rand represents a uniform distribution arbitrary value within $[0, 1]$. Ψ_d has an impact on the search space of the CAS approach. When Ψ_d is smaller, the search space would be larger. when the value of Ψ_d is larger, the search space would be smaller. The search space is fixed to be $[-w_d/2, w_d/2]$, and $w_d \approx 7.5/\Psi_d$.

3 Experimental Validation

The experimental result analysis of the CAS-WELM technique is validated using benchmark dataset. The initial dataset is named as ISOT Fake News Dataset [19] (Sample Set-1), comprising 44,898 articles (21,417 instances under truthful articles and 23,481 under fake articles). The second Kaggle dataset [20] (Sample Set-2) includes 20,386 articles employed to train the dataset and 5,126 articles are applied to test the dataset. The third dataset [21] (sample set-3) comprises 3,352 articles, both fake and true. The final dataset (Sample Set-4) includes the combination of the dataset.

Tab. 1 and Fig. 3 demonstrate the accuracy analysis of the CAS-WELM technique with other ones [22]. The results indicated that the k-nearest neighbor (KNN) model has attained worse classification results than the other methods. In addition, the logistic regression (LR) model has obtained slightly improved classification performance over the KNN model. Moreover, the Localized Support Vector Machine (LSVM), MLP, and Bagging-decision tree (DT) model has accomplished moderately increased outcomes. Though the random forest (RF) model has resulted in competitive outcome, the CAS-WELM technique has outperformed the other methods with the higher accuracy of 99.46%, 96.32%, 96.58%, and 94.89% on the test sample sets 1–4 respectively.

Table 1: Accuracy analysis of CAS-WELM technique with existing methods on the test sample sets 1–4

	Accuracy (%)			
Methods	Sample Set - 1	Sample Set - 2	Sample Set - 3	Sample Set - 4
LR Model	97.00	91.00	91.00	87.00
LSVM Model	98.00	37.00	53.00	86.00
MLP Model	98.00	35.00	94.00	90.00
KNN Model	88.00	28.00	82.00	77.00
RF Model	99.00	35.00	95.00	91.00
Bagging-DT	98.00	94.00	94.00	90.00
CAS-WELM	99.46	96.32	96.58	94.89

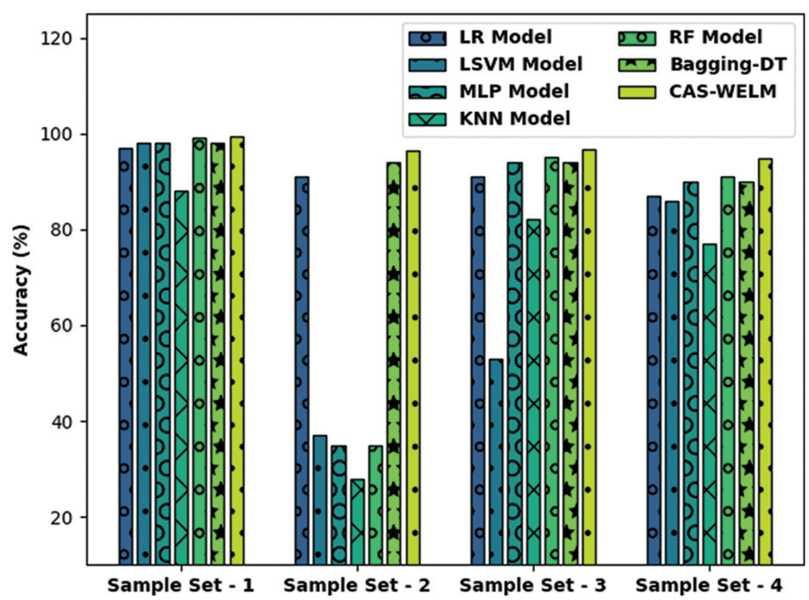

Figure 3: Accuracy analysis of CAS-WELM technique with existing approaches

Tab. 2 and Fig. 4 illustrate the precision analysis of the CAS-WELM approach with other ones. The results indicated that the KNN technique has attained least classification outcomes over the other methods. Besides, the LR approach has reached somewhat higher classification performance over the KNN technique. Moreover, the LSVM, MLP, and Bagging-DT methodology have accomplished moderately increased outcomes. Then, the RF system has resulted in competitive outcome, the CAS-WELM system has demonstrated the other methods with the superior precision of 99.61%, 95.74%, 99.24%, and 95.35% on the test sample sets 1–4 correspondingly.

Table 2: Precision analysis of CAS-WELM technique with existing methods on the test sample sets 1–4

Precision (%)				
Methods	Sample Set - 1	Sample Set - 2	Sample Set - 3	Sample Set - 4
LR Model	98.00	92.00	93.00	88.00
LSVM Model	98.00	31.00	54.00	88.00
MLP Model	97.00	32.00	93.00	92.00
KNN Model	91.00	22.00	85.00	80.00
RF Model	99.00	30.00	98.00	92.00
Bagging-DT	98.00	94.00	93.00	90.00
CAS-WELM	99.61	95.74	99.24	95.35

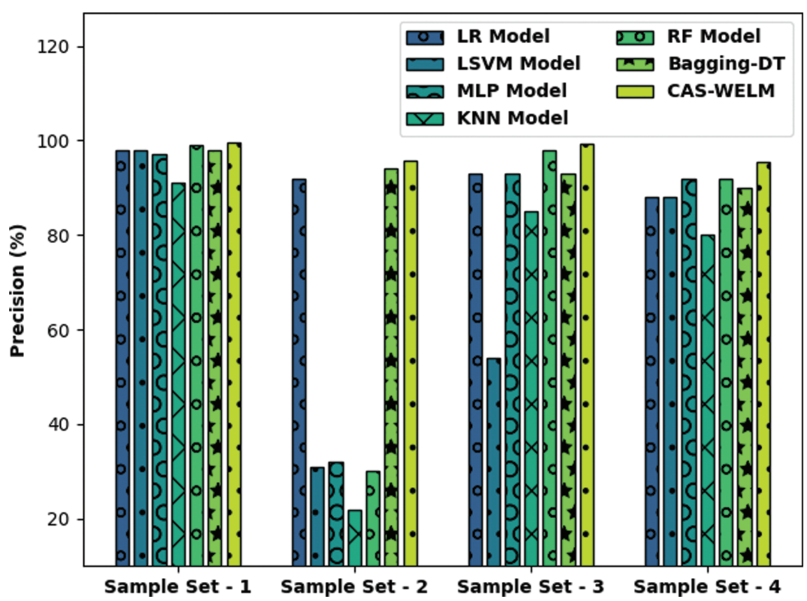

Figure 4: Precision analysis of CAS-WELM technique with existing approaches

Tab. 3 and Fig. 5 showcases the recall analysis of the CAS-WELM approach with other ones. The outcomes referred that the KNN algorithm has gained poor classification results over the other methods. Similarly, the LR technique has obtained slightly enhanced classification performance over the KNN technique. Likewise, the LSVM, MLP, and Bagging-DT approach has accomplished moderately increased outcomes. Eventually, the RF system has resulted in competitive outcome, the CAS-WELM method has exhibited the other techniques with the maximal recall of 100%, 98.24%, 100%, and 95.84% on the test sample sets 1–4 correspondingly.

Table 3: Recall analysis of CAS-WELM technique with existing methods on the test sample sets 1–4

	Recall (%)			
Methods	Sample Set - 1	Sample Set - 2	Sample Set - 3	Sample Set - 4
LR Model	98.00	90.00	92.00	86.00
LSVM Model	98.00	32.00	100.00	86.00
MLP Model	100.00	36.00	96.00	88.00
KNN Model	87.00	24.00	81.00	74.00
RF Model	100.00	34.00	93.00	91.00
Bagging-DT	97.00	95.00	94.00	91.00
CAS-WELM	100.00	98.24	100.00	95.84

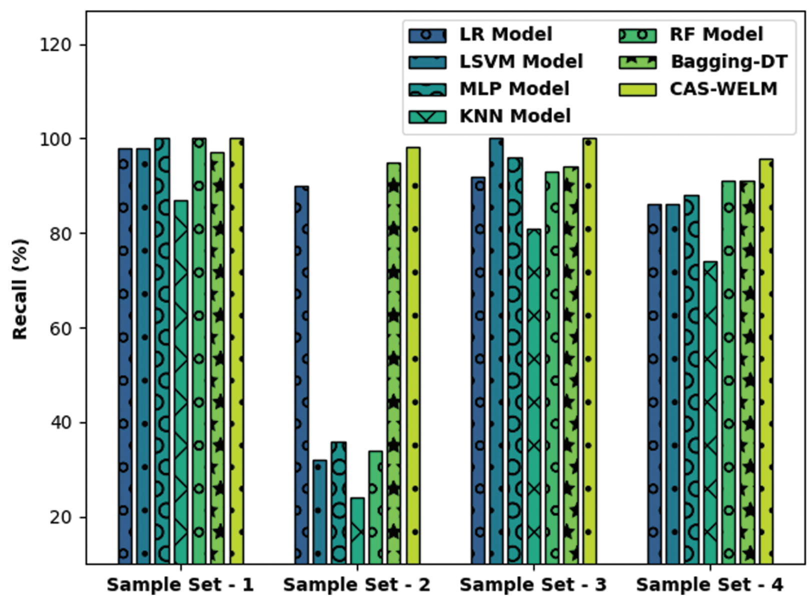

Figure 5: Recall analysis of CAS-WELM technique with existing approaches on test sample sets 1–4

Tab. 4 and Fig. 6 illustrates the F-score analysis of the CAS-WELM technique with other ones. The results show that the KNN method has gained minimal classification outcomes over the other approaches. Besides, the LR technique has obtained somewhat enhanced classification performance over the KNN technique. Moreover, the LSVM, MLP, and Bagging-DT approach has accomplished moderately higher outcomes. At last, the RF system has resulted in competitive outcome, the CAS-WELM technique has outperformed the other methods with the increased F-score of 99.36%, 96.48%, 98.88%, and 96.23% on the test sample sets 1–4 correspondingly.

Table 4: F-score analysis of CAS-WELM technique with existing methods on the test sample sets 1–4

Precision (%)				
Methods	Sample Set - 1	Sample Set - 2	Sample Set - 3	Sample Set - 4
LR Model	98.00	91.00	92.00	87.00
LSVM Model	98.00	32.00	70.00	87.00
MLP Model	98.00	34.00	95.00	90.00
KNN Model	89.00	23.00	83.00	77.00
RF Model	99.00	32.00	95.00	91.00
Bagging-DT	98.00	94.00	94.00	90.00
CAS-WELM	99.36	96.48	98.88	96.23

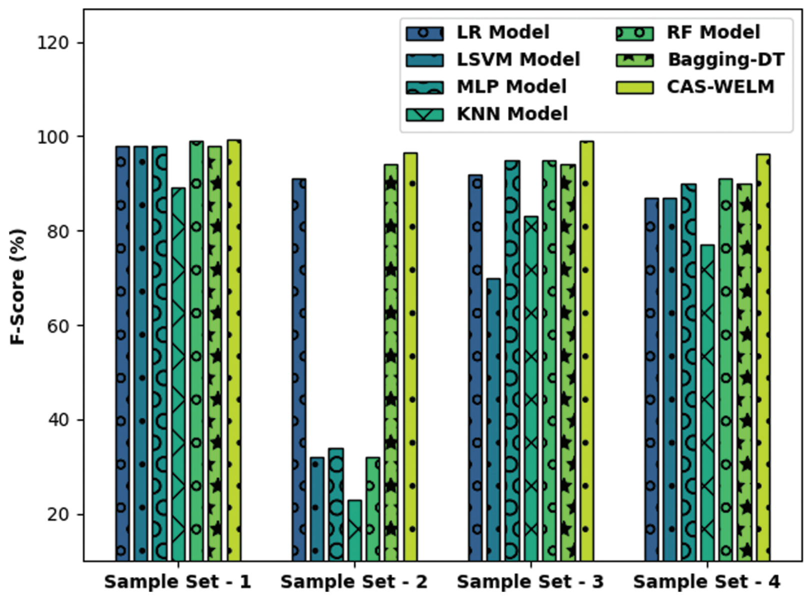

Figure 6: F-score analysis of CAS-WELM technique with existing approaches on test sample sets 1–4

Fig. 7 demonstrates the accuracy and loss graph analysis of the CAS-WELM technique on the test sample sets 1 and 2. The results show that the accuracy value tends to increase and loss value tends to decrease with an increase in epoch count. It is also observed that the training loss is low and validation accuracy is high on test sample sets 1 and 2.

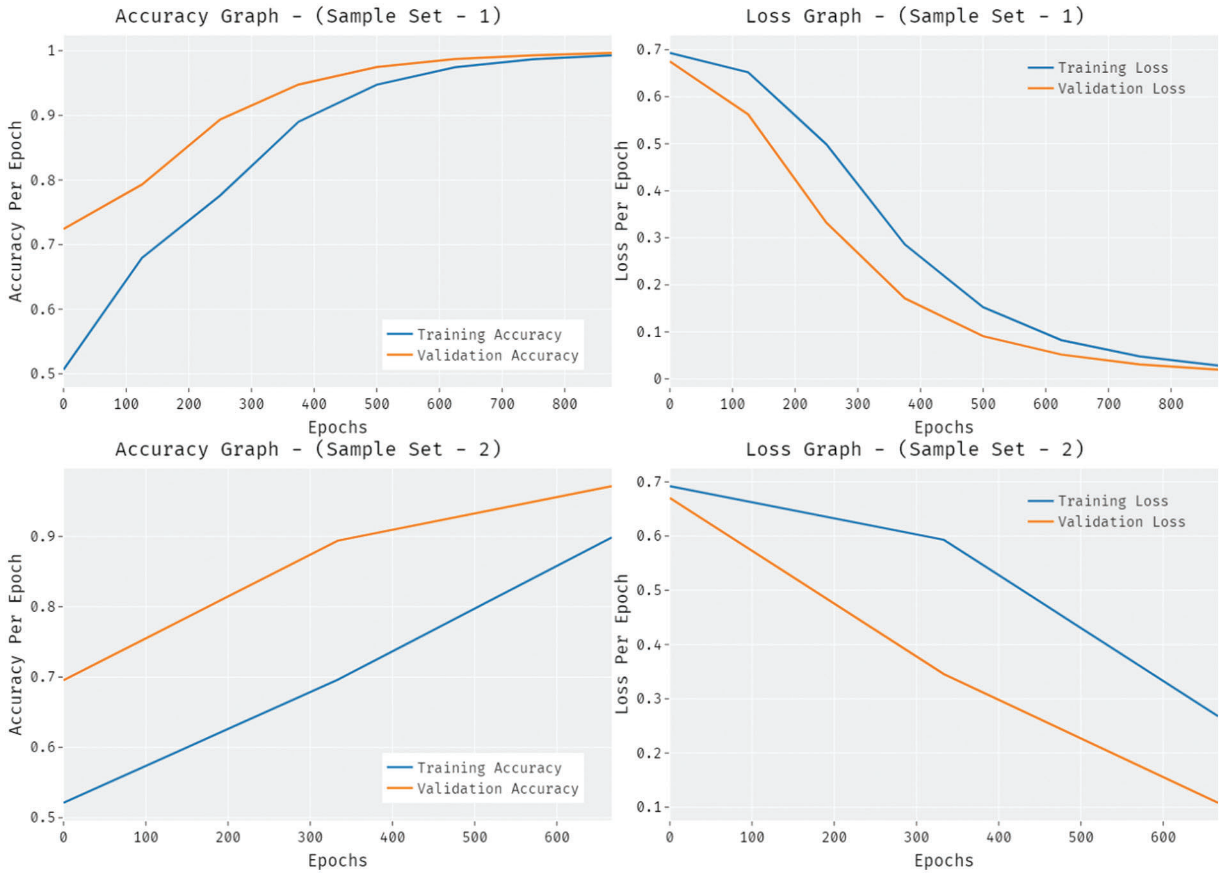

Figure 7: Accuracy and loss analysis of CAS-WELM technique under test sample sets 1 and 2

Fig. 8 offers the accuracy and loss graph analysis of the CAS-WELM methodology on the test sample sets 3 and 4. The outcomes demonstrated that the accuracy value tends to be higher and loss value tends to lower with higher epoch count. It is also experiential that the training loss is minimum and validation accuracy is high on the test sample sets 3 and 4.

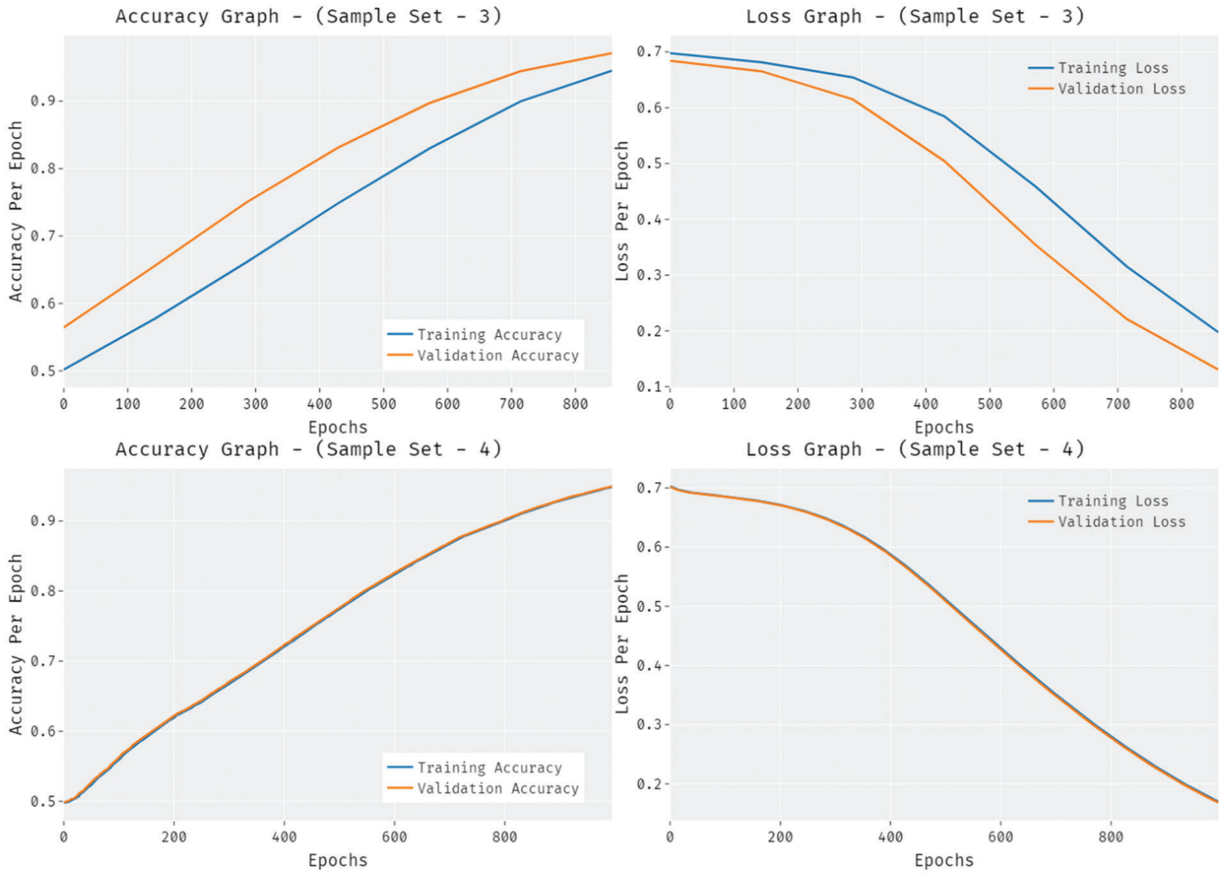

Figure 8: Accuracy and loss analysis of CAS-WELM technique under test sample sets 3 and 4

4 Conclusion

In this study, a novel CAS-WELM technique has been developed for Cybersecurity Fake News Detection and Classification. The CAS-WELM technique mainly intends to discriminate news into fake and real. The CAS-WELM technique undergoes different stages of operations namely pre-processing, Glove based word embedding, N-gram based feature extraction, WELM based classification, and CAS based parameter optimization. Besides, the weight value of the WELM model can be optimally adjusted by the use of CAS algorithm. The performance validation of the CAS-WELM technique is carried out using the benchmark dataset and the results are inspected under several dimensions. The experimental results reported the enhanced outcomes of the CAS-WELM technique over the recent approaches. In the future, advanced deep learning models can be utilized to classify and detect fake news in social networking platform.

Acknowledgement: The authors deeply acknowledge the Researchers supporting program (TUMA-Project-2021-27) Almaarefa University, Riyadh, Saudi Arabia for supporting steps of this work. The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

Funding Statement: This research was supported by the Researchers Supporting Program (TUMA-Project-2021-27) Almaarefa University, Riyadh, Saudi Arabia. Taif University Researchers Supporting Project number (TURSP-2020/161), Taif University, Taif, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero and C. R. Jiménez, "The impact of term fake news on the scientific community scientific performance and mapping in web of science," *Social Sciences*, vol. 9, no. 5, pp. 73, 2020.

[2] T. Hopp, "Fake news self-efficacy, fake news identification, and content sharing on Facebook," *Journal of Information Technology & Politics*, pp. 1–24, 2021. https://doi.org/10.1080/19331681.2021.1962778.

[3] N. K. Conroy, V. L. Rubin and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the American Society for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[4] A. Robb, "Anatomy of a fake news scandal," *Rolling Stone*, vol. 1301, pp. 28–33, 2017.

[5] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 Election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, 2017.

[6] V. Rubin, N. Conroy, Y. Chen and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proc. of the Second Workshop on Computational Approaches to Deception Detection*, San Diego, California, pp. 7–17, 2016.

[7] K. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[8] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, pp. 1–23, 2020.

[9] A. Roy, K. Basak, A. Ekbal and P. Bhattacharyya, "A deep ensemble framework for fake news detection and classification," arXiv preprint arXiv: 1811.04670, 2018.

[10] N. Aslam, I. U. Khan, F. Alotaibi, L. Aldaej and A. Aldubaikil, "Fake Detect: A deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, no. 4, pp. 1–8, 2021.

[11] A. Agarwal, M. Mittal, A. Pathak and L. M. Goyal, "Fake news detection using a blend of neural networks: An application of deep learning," *SN Computer Science*, vol. 1, no. 3, pp. 143, 2020.

[12] Z. Khanam, B. N. Alwasel, H. Siraf and M. Rashid, "Fake news detection using machine learning approaches," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1, pp. 012040, 2021.

[13] W. H. Bangyal, R. Qasim, N. U. Rehman, Z. Ahmad, H. Dar *et al.,* "Detection of fake news text classification on covid-19 using deep learning approaches," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 12, pp. 1–14, 2021.

[14] T. Chauhan and H. Palivela, "Optimization and improvement of fake news detection using deep learning approaches for societal benefit," *International Journal of Information Management Data Insights*, vol. 1, no. 2, pp. 100051, 2021.

[15] C. Zhou, C. Sun, Z. Liu and F. Lau, "A C-LSTM neural network for text classification," arXiv preprint arXiv: 1511.08630, 2015.

[16] S. Ding, H. Zhao, Y. Zhang, X. Xu and R. Nie, "Extreme learning machine: Algorithm, theory and applications," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 103–115, 2015.

[17] Z. Xu, J. Liu, X. Luo, Z. Yang, Y. Zhang *et al.,* "Software defect prediction based on kernel PCA and weighted extreme learning machine," *Information and Software Technology*, vol. 106, no. 6, pp. 182–200, 2019.

[18] M. Wan, C. Wang, L. Li and Y. Yang, "Chaotic ant swarm approach for data clustering," *Applied Soft Computing*, vol. 12, no. 8, pp. 2387–2393, 2012.

[19] H. Ahmed, I. Traore and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, pp. e9, 2018.

[20] Kaggle, Fake News, Kaggle. San Francisco, CA, USA, 2018. [Online]. Available: https://www.kaggle.com/c/fake-news.

[21] Kaggle, Fake News Detection, Kaggle. San Francisco, CA, USA, 2018. [Online]. Available: https://www.kaggle.com/jruvika/fake-news-detection.

[22] I. Ahmad, M. Yousaf, S. Yousaf and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, no. 5, pp. 1–11, 2020.