SCIENTIFIC CULTURE

www.sci-cult.com

*SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 2478-2491*
*Open Access. Online & Print*

**DOI: 10.5281/zenodo. 121126174**

# AN EFFECTIVENESS OF WEBSITE CLASSIFICATION IN WEB MINING USING BIDIRECTIONAL GATED ATTENTION RECURRENT NEURAL NETWORK WITH IMPROVED COATI OPTIMIZATION ALGORITHM

**S.Jaiganesh[1], L.R.Aravind Babu[2], T.Padmapriya[3], Christhu Raj M R[4]**

*[1]Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India. E-mail: jganesh0@gmail.com, ORCID: 0009-0001-0436-3218*
*[2]Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India. E-mail: er.arvee@rediffmail.com, ORCID: 0009-0002-4093-330X*
*[3]Melange Publications, Puducherry, India. E-mail: padmapriyaa85@ptuniv.edu.in, ORCID: 0000-0002-9766-4203*
*[4]Directorate of Learning and Development, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, 603203, India. E-mail: christhm1@srmist.edu.in, ORCID: 0000-0002-6461-6301*

*Received: 28/08/2025*
*Accepted: 12/01/2026*

*Corresponding Author: S.Jaiganesh (jganesh0@gmail.com)*

## ABSTRACT

***With the huge upsurge in the volume of information accessible on the World Wide Web (WWW) nowadays, and the development need for an above method to access this information, there was a powerful resurgence of interest in web mining research. Web mining is a dangerous problem in data mining and other information process methods to determine valuable patterns. With the upsurge in the amount of websites and web users, the necessity for the classification of website increases attraction. The classification of the website based on URLs only plays a vital role, as the contents of web pages are not necessarily attained for classification. Nowadays, Machine learning (ML) and Deep learning (DL) can be significant in finding known and novel malicious URLs. Various types of research are performed on malicious URL classification and detection utilizing different ML techniques. In this manuscript, we present a Website Classification using the Bidirectional Gated Attention Recurrent Neural Network with Improved Coati Optimization (WCBGARNN-ICO) technique. The key intention of the WCBGARNN-ICO method is to identify websites in web mining. At first, the WCBGARNN-ICO model applies the quality of text preprocessing with different levels to attain clear data and extract significant data. For the extraction of the feature process, the bidirectional encoder representations from the transformers (BERT) method can be exploited. Besides, the attention-based bidirectional gated recurrent neural network (A-BGRNN) model is employed for the website classification process. At last, the improved coati optimization algorithm (ICOA) is deployed to fine-tune the hyperparameter of the A-BGRNN technique. The experimental study of the WCBGARNN-ICO algorithm is tested on a benchmark database and the findings are measured with numerous measures. The experimental findings highlighted the development of the WCBGARNN-ICO system over other approaches.***

**KEYWORDS:** Website Classification, Web Mining, Improved Coati Optimization, Bidirectional Gated Attention Recurrent Neural Network, Preprocessing.

## 1. INTRODUCTION

With the fast improvement of the World Wide Web (WWW), an enormous amount of information is available now to web users owing to their publishing freedom, Lower cost, and high availability that have promoted to its fame [1]. In addition, the web pages and multimedia and text modules, contain context features like metadata, hyperlinks, and HTML tags. Classification plays a crucial role in various information management and recovery tasks [2]. On the Web, page content classification can be vital to determined crawling, to the supported growth of web manuals, to topic-specified web link study, and the study of the current web structure [3]. Also, the Classification of web pages could assist in enhancing the web searching quality. Web mining plays an important part in finding valuable pattern logs or information to aid in classification [4]. The web mining technique is a data mining application of an algorithm or technique and knowledge discovery process that concentrates on extracting data patterns instantly from the web data that assist in improving the search engine's ability to identify the contents and predict its user behaviors [5].

Recent webpage classification methods utilize a selection of information to identify a target page: anchor text, its hyperlink structure, the text of the page itself, and the structure of the link from pages directing to the location and its target page (specified by its URL) [6,24]. Of this information, a website uniform resource locator (URL) is the cheapest to attain and one of the more useful resources based on classification. URLs were frequently intended to be easily recollected by humans, and web pages that correspond to great design methods will encode valuable terms, which define their resources in the domain name of the website [7]. Currently, Machine learning (ML) can be significant in recognizing identified and novel malicious URLs. ML is a mechanism by which computers have been trained to understand data, allowing them to autonomously produce decisions or predictions [8]. The utmost frequently utilized ML method for recognizing malicious URLs is classification, a domain of supervised ML. Deep learning (DL) methods are used for an extensive range of tasks in NLP, enhancing language modeling for a more prolonged context [9]. Classification or categorization of text can be regarded as a vital topic in the area of natural language processing (NLP); it is also an important tool in various areas like searching the web, filtering information, and categorization topics [10].

This manuscript presents a Website Classification using the Bidirectional Gated Attention Recurrent Neural Network with Improved Coati Optimization (WCBGARNN-ICO) technique. At first, the WCBGARNN-ICO model applies the quality of text preprocessing with different levels to attain clear data and extract significant data. For the extraction of the feature process, the bidirectional encoder representations from the transformers (BERT) method can be exploited. Besides, the attention-based bidirectional gated recurrent neural network (A-BGRNN) model is employed for the website classification process. At last, the improved coati optimization algorithm (ICOA) is deployed to fine-tune the hyperparameter of the A-BGRNN technique. The experimental findings of the WCBGARNN-ICO algorithm can be tested on a benchmark database and the findings are measured with numerous measures.

## 2. RELATED WORKS

Tavasoli et al. [11] propose the Web-Based Intelligent Packaging Evaluation (WIPE) platform, a new technique to evaluate the implementation of packaging systems and products in the e-commerce distributing sectors. Various conventional techniques, which mainly trust on lab valuations under precise circumstances, WIPE tackles the distinctive challenges modeled by e-commerce distribution, like improved unexpected dangers and handling points, which usual manual tests cannot detain. Rao and Venkatraman [12] proposed a web mining study direction, freely accessible tools, and their various applications to commercial support consumers. The WWW is an enormous, data center for a range of applications. The web has rich and dynamic loads of hyperlink information. It enables Web page access, and use of information and offers various resources for data mining. Afif [13] proposed to study the PSAU website quality based on data mining methods. The initial phase: was gathering feedback about the PSAU website utilizing a review. Afterward, the data mining methods are utilized as predictive and descriptive methods. The descriptive method has been employed to define and extract the main indexes of website quality.

Amane et al. [14] proposed a study that contains 2 major stages. Initially, the author extracts metadata from learning objects, utilizing the method of web search approaches like feature selection (FS) methods that are mostly executed to identify the greatest group of features, which enables us to create valuable methods. The important role of FS in learning object classifiers is to recognize the relevant features and reduce redundant features from an extremely

dimensional dataset. Next, the author recognizes the learning objects based on a specific form of equality utilizing a Multi-Label Classification (MLC) derived from Fuzzy C-Means (FCM) techniques. Zhao et al.[15] present an e-commerce web page recommending a solution, which incorporates BP neural networks (NN) and semantic web (SW) mining. Initially, the user weblog searches were processed, and five features were extracted: time consumption priority, content priority, input deviance quantity, online shopping users' implicit/explicit reviews on the website, and suggestion semantics. Later, these features were utilized as BP-NN input features to identify and classify the importance of the web page concluding output.

Zaidi et al. [16] proposed a detailed study of software technologies, which considers either the content or technical features. A comprehensive study of ontology, ontology creation, interoperability guidelines, and kinds of ontologies, ontology editors, and real-world applications of ontologies in healthcare were specified in this research. The owner of the website needs to assist visitants by providing valuable information, which will assist them in attaining their goals. SheykhAbbasi et al.[17] introduced a method for extracting user interests and web personalization utilizing a hybrid technique of web content and web-usage mining. Thus, the directional forms of web users and the interests of every user on web pages of a Persian website are mined over topic modeling and web-usage mining, correspondingly. Then, users are grouped utilizing the dependence distributing method, and 25 classes are extracted. For understanding the behavioral patterns of the web user, they are classified utilizing the SVM method based on the interest of the user and directional behaviors.

## 3. PROPOSED METHODOLOY

In this manuscript, we have presented a novel WCBGARNN-ICO technique. The key intention of the WCBGARNN-ICO algorithm is to classify websites in web mining. To achieve that, the WCBGARNN-ICO model has text preprocessing, BERT feature extractor, A-BGRNN-based classification, and ICOA-based parameter fine-tuning processes. Fig. 1 represents the entire flow of the WCBGARNN-ICO algorithm.

### *3.1. Pre-processing*

At first, the WCBGARNN-ICO model applies the quality of text preprocessing with different levels to attain clear data and extract significant data. The cleaning process performed includes accounting for the English language and semantic-specific characteristics to eliminate stop words and lemmatization tokens (words) [18]. Additionally deliberates on features of language transcribed on social media (slang, emoticons, informal, and so on). The value of text pre-processing is important to gain clear data and remove significant data. A Python library keen to pre-process tweets has been applied in the primary position to eliminate hashtags, URLs, smileys, emojis, and reserved words (RT for retweets). Nevertheless, the data attained was not considered clear and adequate. So, the next main stages have been performed additionally:

Removal of HTML Tags: Regarding web-scraped data or content from HTML sources, it's important to remove HTML tags to remove important text. This stage guarantees that just the related content remains to be studied.

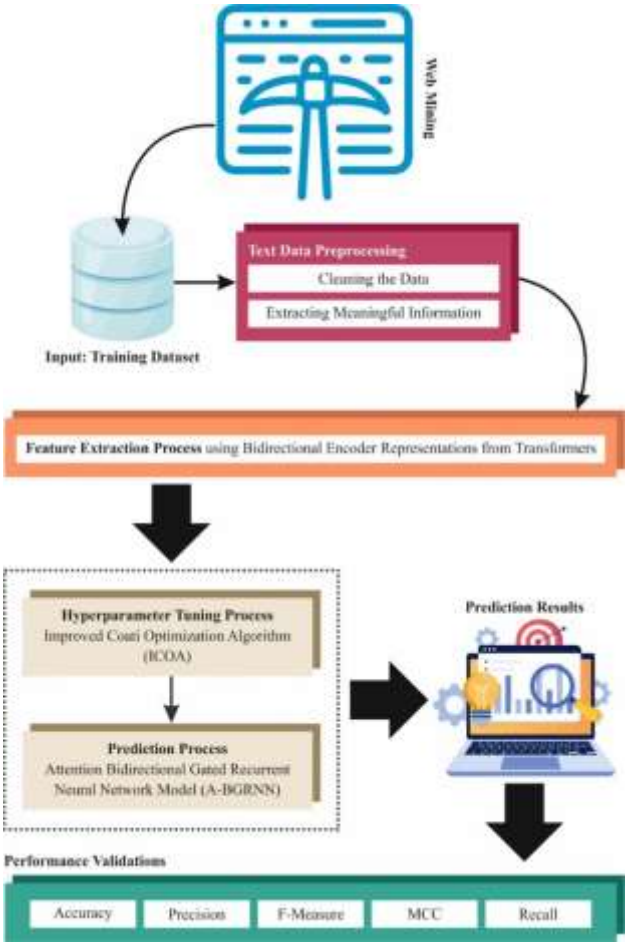


*Figure 1: Overall Flow of the WCBGARNN-ICO Technique.*

**Removal of punctuation symbols**: Punctuation symbols are superfluous in text data study and it's usual to remove them. There are some Python libraries for the Removal of punctuation symbols;

these symbols also can removed and detected with standard expressions. The Python library, called string, has been applied to remove punctuation symbols along with special characters (e.g., #, *, |, (,), and @).

**Removal of Numbers**: Numeric values can present unrelated data, particularly in contexts where numbers don't deliver important meaning. Eliminating them can lessen the text dimensionalities and increase the model's concentration on important words.

**Eliminating Extra Spaces**: Text data can occasionally comprise extra spaces, which include no value to the study. Eliminating these guarantees clear and constant formatting.

**Remove Stop Words:** Stop words occur in every sentence in the text dataset and they don't include meaning to the task of classification. They contain no meaning. It is essential to utilize a particular list of stop words for a single project possibly. Particular language-specific text data classification and list of stop words must be changed. In the English language, stop words correspond to me, myself, for, i, their, of, his, some, him, that, can, such, re, not no, ve, nor, will, and so on. In applications of sentiment analysis, investigators do not remove some positive or negative words as they are valuable to characterize the user's emotions or reviews. Less, No, Like Not, and so forth.

**Lowercasing**: Even though lowercasing is unnoticed by the investigators, it's the simpler and most effective manner to pre-process a text owing to its aids in condensing the text. Capitalization has been applied to mark the start of the sentence in some type of document. It appears that there is no change between lowercase and uppercase words; however, in a document with numerous sentences, capitalization is a significant difficulty if the text is classified. One of the most common approaches to perform capitalization is transforming each of the letters into lowercase. Before classification, all uppercase letters are usually transformed into lowercase letters. This technology predicts each of the words in the text file to contain the same feature. Lowercasing has been used typically in text mining and NLP subjects and is particularly cooperative with the reliability of predictable output.

**Stemming:** Stemming includes reducing words to their root or base form by splitting off suffixes or prefixes. These assistances in combining related words in a regular form, therefore decreasing variabilities within the text.

**Lemmatization words:** the objective is to transform words into their root form, as the position tag and context. It is the chosen model for stemming, the last is simpler but with low execution outcomes as the suffixes of the word are shortened.

### *3.2. BERT Feature Extractor*

For the feature process extraction, the BERT method can be exploited. It is according to the article "attention is all you need", depends on the attention mechanism to understand the symbol and text of sentences [19]. It is licensed by a team of researchers and investigators from the firm Google, and nowadays a larger amount of new NLP applications, that contain translation, search, and speech recognition, are originated from it.

The mechanical basis and structural design of BERT depend on transformers, which work based on the method of attention, which provides it the capability to change the concentration on various sections of input data in the processing method itself. It gives the knowledge of the meaning of words in the sentence. Transformers are based on numerous headings of attention, hence they handle portions of the sentence in parallel, which improves the efficiency and speed of management.

It contains the capability to effectively handle natural language in both ways. It is critical for a deeper consideration of the same language. Namely innovative because earlier methods depended on processing language in some direction. Utilizing the masked language model (MLM) theory, training occurs by words hiding at random in a sentence and then predicting the unseen words according to the content. Each of these assistances him in learning for understanding structures and language patterns.

BERT can utilize the transfer learning theory, it is trained on a larger dataset and then attuned to specific tasks. Its application in the NLP domain is enormous, from text summarization, question answering, text classification, and a few others. Due to all of this, nowadays BERT has provided within best solution for various tasks of this kind. By its capability to interpret and understand, it unlocked a novel section in the area of intelligent systems that can understand and communicate human language.

### *3.3. Classification using the A-BGRNN Model*

Besides, the A-BGRNN model is employed for the website classification process. Owing to the complex framework of the memory unit of Long Short-Term Memory (LSTM), there are longer training time problems [20]. They presented another version of LSTM-GRU (Gated Recurrent Unit) [26]. The memory unit of GRU associations the $f$ forgetting gate and the $i$ input gate within the LSTM toward $z$

upgrade gate, which in addition keeps significant features, apart from resolving the problem of longer dependency, in the meantime the architecture is easier in comparison to LSTM. By time $t$, for an assumed input $X_t$, the hidden layer (HL) of the outputs of GRU $h_t$, the process of computation is as demonstrated

$$z_t = \sigma(W_Z \cdot [h_{t-1}, x_t]) \quad (1)$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$
$$\tilde{h}_t = \tanh(W \cdot [\gamma_t * h_{t-1}, x_t]) \quad (3)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

Here, $W$ denotes the weighted matrix relating the dual layers, and tanh and $\sigma$ represent the activation function. $z$ and $r$ signifies the update and the reset gates correspondingly.

In contrast to problems of sequence, the normal RNN employs the prior informations depending on the forward input series, however isn't deliberate the subsequent information. Concerning these problems, the Bi-directional RNN (BRNN) method presented, while learning the above-mentioned information, also learns the subsequent data. The important concept is to utilize dual RNN to handle the sequences of forward as well as backward, correspondingly. The output is further associated with a similar output layer as bi-directional contextual information for the sequence of features is noted down. According to the BRNN, the BGRU algorithm is gained by substituting the HL neurons within the BRNN through the memory unit of GRU.

For an assumed $n$-dimensional input $(x_1, x_2, \ldots, x_n)$. According to time $t$, the HL of BGRU outputs $h_t$. The mathematical computation is as represented:

$$\overrightarrow{h_t} = \sigma\left(W_{x\vec{h}} x_t + W_{\vec{h}\vec{h}} \vec{h}_{t-1} + b_{\vec{h}}\right) \quad (5)$$
$$\overleftarrow{h_t} = \sigma\left(W_{x\overleftarrow{h}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}\right) \quad (6)$$
$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \quad (7)$$

Now, $W$ represents the weighted matrices joining the dual layers, $b$ denotes the biased vector, $\sigma$ signifies activation function, $\overleftarrow{h_t}$ and $\overrightarrow{h_t}$ represents outputs of negative and positive GRU correspondingly. $\oplus$ is element-to-element sum. Fig. 2 illustrates the structure of A-BGRNN.

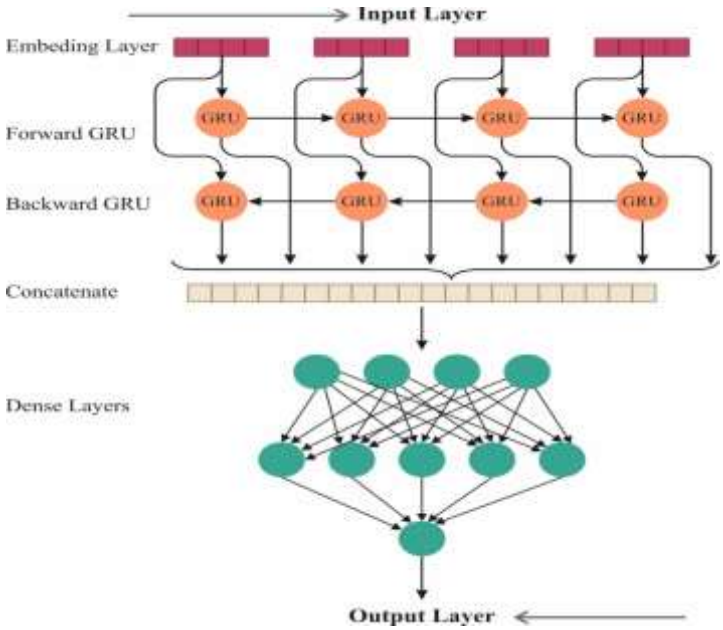


***Figure 2: Structure of A-BGRNN Model.***

During the text examination task, the AM was applied to denote the connection among the words in the sentence of the text and the resultant output. This algorithm is primarily utilized for the machine conversion task. The feed-forward AM accepted in this work is a straightforward simplification of the traditional AM. The generalization model is to build a solitary vector C from the complete series, made in the following

$$e_t = a(h_t) \quad (8)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \quad (9)$$

$$c = \sum_{t=1}^{T} \alpha_t h_t \quad (10)$$

Here, $a$ represents the learning function, at present it's only decided by $h_t$. Among the above-mentioned equation, the AM is measured as building a stable length of the layer of embedding $\mathcal{C}$ of the input series by computing an adaptable weight

average of the series of states $h$.

We gain the last sentence-pair depiction applied for classifications:

$$h^* = \tanh(c) \quad (11)$$

### 3.4. ICOA-based Parameter Tuning

At last, the ICOA is deployed to fine-tune the hyperparameter of the A-BGRNN technique. COA is a new meta-heuristic model $M$ that simulates coati actions in nature [21,25]. The COA basic is to imitate the normal behavior of dual coatis like exploration (fleeing from predators) and exploration (attacking and hunting iguanas) dual behaviors. The main phases of COA are presented in the next subgroup.

During this initialization phase of the COA, the location of the coatis in searching space can be produced at random for the COA by utilizing the representation in Eq. (12)

$$X_i: x_{i,j} = b_j^L + rand(0,1) \cdot \left(b_j^U - b_j^L\right), i = 1,2,\dots,N, j = 1,2,\dots,m \quad (12)$$

In Eq. (12): $X_i$ refers to the location within the search space from the $i_{th}$ coati, $x_{i,j}$ symbolizes the $j_{th}$ decision variable value. $b_j^U$ and $b_j^L$ denotes the lower and upper limit of the decision variables, correspondingly. $N$ represents several coatis, $m$ signifies decision variable counts.

Stage 1: Attacking and hunting tactic on iguana (stage of exploration)

The initial stage of updating the population of coatis within the searching space is demonstrated depending upon pretending their strategies once attack iguanas. During this tactic, a collection of coatis ascents the tree to approach an iguana and frighten it. Some another coatis stay in a tree, waiting for the iguana to drop to the ground. Once the iguana descends to the ground, the coatis hunt and attack it. This tactic precedents Coatis to travel to various locations within the searching space, it represents the exploration capability of COA's in global search with the problem-solving space. In the stage of exploration, the location of Coati's updated tactic mostly pretends Coati's attacking and hunting behavior of iguanas. The Coati's behavior is separated into 2 stages to overall the attacking and hunting of the iguana. i) Fear. A collection of Coatis climbs a tree to reach an iguana and frighten it. ii) Some other Coatis remain under a tree till the scared iguana drops to the ground, and once the iguana falls, finish the hunt and attack it. These strategies reason Coati to travel to various places in the searching space, successively displaying that the COA optimizer method has the exploration ability to solve the problem space global search. During this COA model, it's assuming the position of the top follower of the population is that of the iguana. It is additionally assuming the Coati counts implementation phases (1) and (2) in every half of the total Coati counts. Therefore, the mathematical formulations of location are:

$$X_i^{P1}: x_{i,j}^{P1} = x_{i,j} + rand(0,1)\left(G_j - Ix_{i,j}\right), i = 1,2,\dots,\left[\frac{N}{2}\right]; j = 1,2,\dots,m \quad (13)$$

Here, $X_{i,j}^{P1}$ represents the original location of the $ith$ Coati in the $jth$ size; $r$ denotes a randomly generated number among [zero and one]; $G_j$signifies iguana location in the $jth$ size, which stands for the location of the top member; $I$ refers to the number selected at random from the set [1,2]; $N$ represents Coati count; $[N/2]$ means main integer not greater $[N/2]$; $m$ stands for decision variable counts.

Afterward, the iguana drops to the ground, it's positioned in an arbitrary place within the searching space. According to this arbitrary position, the Coati by the ground passes over the searching space. These phases can be mimicked by dual equations.

$$G^g: G_j^g = b_j^L + rand(0,1)\left(b_j^U - b_j^L\right) \quad (14)$$

Whereas: $G_j^g$ refers to the iguana's location on the ground in the $j\ th$ size.

$$X_i^{P1}: x_{i,j}^{P1} = \begin{cases} x_{i,j} + rand(0,1) \cdot \left(G_j^g - Ix_{i,j}\right), & F_{G,j}^g \le F_{i,j} \\ x_{i,j} + rand(0,1) \cdot \left(x_{i,j} - G_j^g\right), & other \end{cases} \quad (15)$$

$$i = [N/2] + 1, [N/2] + 2, \cdots, N, \text{and } j = 1,2,\dots,m$$

Here: $F_{G,j}^g$ stands for a value of the objective function of the $jth$ size iguana once it drops to the ground; $F_{i,j}$ represents the value of the objective function of the $ith$ Coati in $jth$ size. When the upgraded individual is superior, the present individual has been upgraded. Or else let that as it stands.

$$X_i = \begin{cases} X_i^{P1}, F_i^{P1} \le F_i \\ X_i, other \end{cases} \quad (16)$$

Now:$F_i^{P1}$ indicates the value of the objective function of the $ith$ Coati on the original location; $F_i$ means the value of the objective function of the $ith$ Coati by the earlier location.

Stage 2: The process of escaping from predators (exploitation stage)

During this stage of exploitation, Coati's position-upgrading tactic mostly imitates the natural behavior of Coati after meeting predators and escaping from predators.

In the exploitation stage, once a hunter assaults Coati, Coati loses its location. Coati's change in this tactic led to its nature in a safer place near its present

location. This determines the COA exploitation systems in local search. To pretend these actions, an arbitrary position has been made close to the position of every Coati depending on the subsequent Eq. (17)

$$b_{j,L}^{loc} = \frac{b_j^L}{t}, b_{j,U}^{loc} = \frac{b_j^U}{t}, t = 1,2,\cdots,T \quad (17)$$

Here: $b_{j,L}^{loc}$ and $b_{j,U}^{loc}$ refers to a local lower and upper limit of the $jth$ decision variable, and $t$ denotes iteration counts; $T$ indicates maximal iteration counts.

$$\begin{cases} X_{i,j}^{P2} = x_{i,j} + (1-2r)\left(b_{j,L}^{loc} + r\left(b_{j,U}^{loc} - b_{j,L}^{loc}\right)\right) \\ i = 1,2,\cdots,N \end{cases} \quad (18)$$

Now:$X_{i,j}^{P2}$ means the original location of the $ith$ Coati inside the $jth$ size. When the upgraded individual is superior, upgrade the present individual, or else let that as it stands.

$$X_i = \begin{cases} X_i^{P2}, F_i^{P2} \le F_i \\ X_i, \quad other \end{cases} \quad (19)$$

Since the individual locations of the new COA are made at random the population range is possible to be destroyed, and it is distribution uniformly within the solution space can't assured, this means the model simply drops into local optimizer. The constant population could accelerate convergence. Hence, it's essential to increase the initialization process of the method. Chaotic mapping (CM) is stochastic and ergodic. When the CM function has been applied to make a chaotic series as the primary location of individual population to create the distribution of population more identical and prevent population uniformities, so increasing the search proficiency. Generally, applied CMs present themselves as a circle map, Chebyshev map, Logistic map, Gauss map, Cubic map, Iterative map, Sine map, Tent map, and Singer map.

The population distribution made by Tent Chaos contains the optimal consistency between the above main chaotic mappings. Hence, this work selects Tent Chaos to increase the distributed qualities of the first population in searching space, and the global searching capability is supported, to increase the solution effectiveness of the model, and Eq. (12) is modified as:

$$X_i: x_{i,j} = b_j^L + \left(b_j^U - b_j^L\right) \cdot z_i (20)$$

The formulation of the tent maps is exposed in Eq. (20), $\alpha = 0.5$.

$$z_{i+1} = \begin{cases} \frac{z_i}{\alpha}, |x_{i,j} \in [0,\alpha) \\ \frac{1-z_i}{1-\alpha}, |x_i \in [\alpha,1) \end{cases} \quad (21)$$

The ICOA improves a fitness function (FF) to achieve superior classification outcomes. It identifies a positive integer to denote the greater outcome of the candidate solutions. In this research, the reduction of the classification rate of error is observed as FF, as presented in Eq. (22)

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{no.of\ misclassified\ samples}{Total\ no.of\ samples} * 100 \quad (22)$$

## 4. RESULT ANAYSIS AND DISCUSSION

The simulated analysis of the WCBGARNN-ICO model is examined by using a website classification database [22]. The database consists of 1376 counts under 14 class labels with various classes as displayed in Table 1. Table 2 denotes the sample tweets.

***Table 1: Details of Dataset***

| Category | Class Labels | No. of Count |
|---|---|---|
| Business/Corporate | C1 | 109 |
| Computers and Technology | C2 | 93 |
| E-Commerce | C3 | 102 |
| Education | C4 | 114 |
| Food | C5 | 92 |
| Games | C6 | 98 |
| Health and Fitness | C7 | 96 |
| Law and Government | C8 | 84 |
| News | C9 | 96 |
| Photography | C10 | 93 |
| Social Networking and Messaging | C11 | 83 |
| Sports | C12 | 104 |
| Streaming Services | C13 | 105 |
| Travel | C14 | 107 |
| Total Number of Count | | 1376 |

***Table 2: Sample Tweets.***

| S. No | Website_url | Cleaned_website_text | Category |
|---|---|---|---|
| 0 | https://www.hotwire.com/ | cheap hotels cars flights minute travel deals hotwire find good deal cheap hotel car flight hotwire save hotel rent car day favorite destination | Travel |
| 1 | https://secure.imvu.com/welcome/ftux/ | imvu 3d avatar social app virtual world virtual reality vr avatars free 3d chat imvu official website imvu 3d avatar social app allow user explore thousand virtual worlds metaverse create 3d avatars enjoy 3d chat meet people world virtual setting spread power friendship | Social Networking and Messaging |
| 2 | https://blazinphoto.com | adrieanablazin photographer specialize distinctive portrait people pet adrieanablazin photography portfolio contact | Photography |
| 3 | https://www.todayonline.com/ | late news singapore world todayonline late trend news singaporeasia world big read commentaries | News |
| 4 | http://www.thegymla.com/ | gym ultimate training facility gym losangeles fit la personal training weight lifting weight loss body building fitness aerobics spin cardio | Health and Fitness |
| 5 | http://www.boardgamegeek.com/game/2545 | contigo board game boardgamegeek line control mancala system board game boardgamesboardgame board game game hobby boardgamegeek geek geekdo | Games |

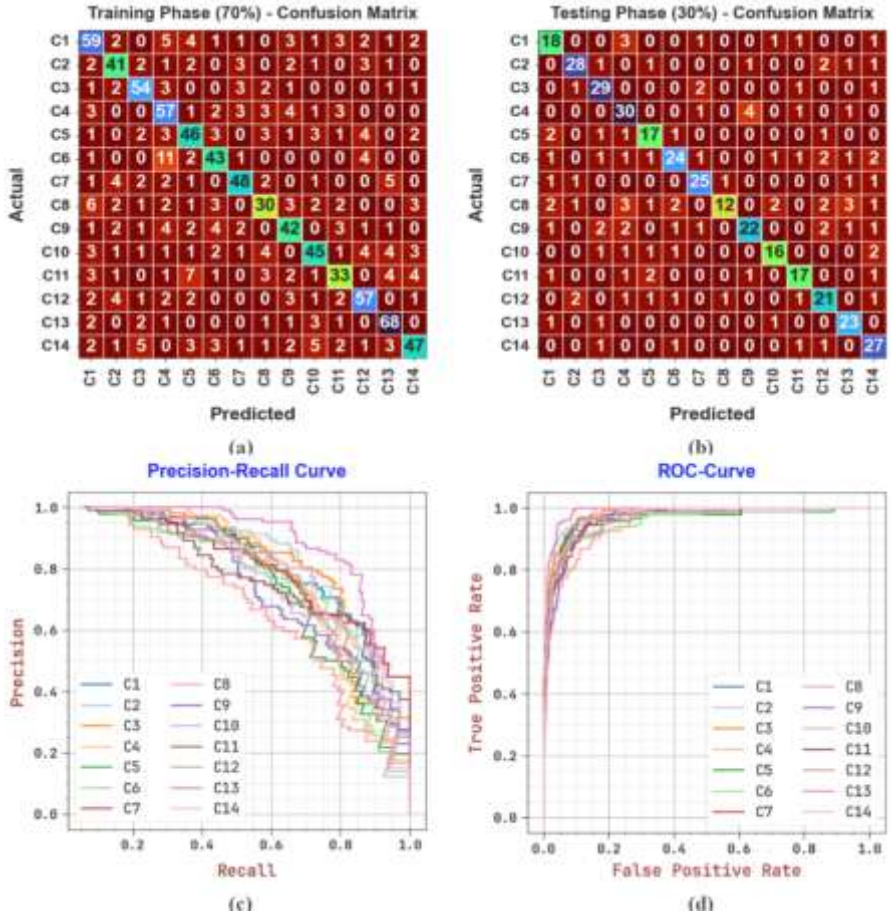


***Figure 3: Classifier Outcomes of (a-b) 70% TRAP and 30% TESP of Confusion Matrices and (c-d) PR and ROC Curves.***

Fig. 3 denotes the classification study of the WCBGARNN-ICO model on the test database. Figs. 3a-3b displays the confusion matrix with precise classification and recognition of all 14 classes on a 70% TRAP: 30% TESP. Fig. 3c shows the study of PR, representing greater values across all 14 class labels. Eventually, Fig. 3d demonstrates the study of ROC, validating efficient values with higher ROC values for different classes.

Table 3 signifies website classification findings of the

WCBGARNN-ICO system under 70 %TRAP and 30% TESP.

In Fig. 4, the average findings represented by the WCBGARNN-ICO system on 70% TRAP are underlined. The findings revealed that the WCBGARNN-ICO technique got effective website classification. On 70%TRAP, the WCBGARNN-ICO method attains average $accu_y$ of 95.65%, $prec_n$ of 69.41%, $reca_l$ of 69.04%, $F_{measure}$ of 69.06%, and MCC of 66.82%.

Fig. 5 offers average values of the WCBGARNN-ICO model under 30%TESP. The findings displayed that the WCBGARNN-ICO system obtained effective website classification. On 30%TESP, the WCBGARNN-ICO model attains average $accu_y$ of 96.40%, $prec_n$ of 75.29%, $reca_l$ of 74.64%, $F_{measure}$ of 74.17%, and MCC of 72.68%.

***Table 3: Website Classification Outcome of WCBGARNN-ICO Model under 70%TRAP and 30%TESP.***

| Class Labels | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{measure}$ | $MCC$ |
|---|---|---|---|---|---|
| **TRAP (70%)** | | | | | |
| C1 | 94.50 | 67.82 | 70.24 | 69.01 | 66.00 |
| C2 | 96.26 | 68.33 | 70.69 | 69.49 | 67.51 |
| C3 | 96.78 | 76.06 | 79.41 | 77.70 | 75.98 |
| C4 | 94.18 | 61.29 | 74.03 | 67.06 | 64.24 |
| C5 | 94.91 | 63.89 | 66.67 | 65.25 | 62.52 |
| C6 | 96.05 | 69.35 | 69.35 | 69.35 | 67.25 |
| C7 | 96.57 | 76.19 | 72.73 | 74.42 | 72.61 |
| C8 | 95.43 | 61.22 | 54.55 | 57.69 | 55.39 |
| C9 | 95.53 | 65.62 | 66.67 | 66.14 | 63.75 |
| C10 | 95.33 | 70.31 | 63.38 | 66.67 | 64.26 |
| C11 | 95.33 | 64.71 | 55.00 | 59.46 | 57.21 |
| C12 | 96.16 | 75.00 | 76.00 | 75.50 | 73.41 |
| C13 | 96.78 | 77.27 | 86.08 | 81.44 | 79.82 |
| C14 | 95.33 | 74.60 | 61.84 | 67.63 | 65.46 |
| **Average** | **95.65** | **69.41** | **69.04** | **69.06** | **66.82** |
| **TESP (30%)** | | | | | |
| C1 | 96.13 | 66.67 | 72.00 | 69.23 | 67.22 |
| C2 | 97.09 | 84.85 | 80.00 | 82.35 | 80.81 |
| C3 | 96.85 | 78.38 | 85.29 | 81.69 | 80.06 |
| C4 | 95.16 | 69.77 | 81.08 | 75.00 | 72.59 |
| C5 | 96.85 | 70.83 | 73.91 | 72.34 | 70.69 |
| C6 | 95.40 | 77.42 | 66.67 | 71.64 | 69.38 |
| C7 | 97.34 | 80.65 | 83.33 | 81.97 | 80.54 |
| C8 | 95.40 | 85.71 | 41.38 | 55.81 | 57.69 |
| C9 | 95.88 | 78.57 | 66.67 | 72.13 | 70.20 |
| C10 | 97.34 | 76.19 | 72.73 | 74.42 | 73.04 |
| C11 | 97.09 | 73.91 | 73.91 | 73.91 | 72.37 |
| C12 | 95.64 | 67.74 | 72.41 | 70.00 | 67.70 |
| C13 | 97.34 | 74.19 | 88.46 | 80.70 | 79.64 |
| C14 | 96.13 | 69.23 | 87.10 | 77.14 | 75.65 |
| Average | 96.40 | 75.29 | 74.64 | 74.17 | 72.68 |

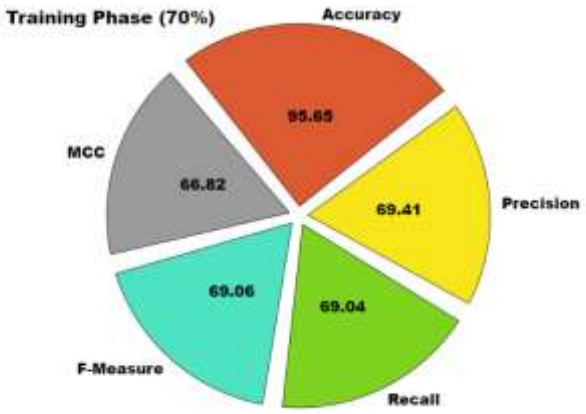


***Figure 4: Average Outcome of WCBGARNN-ICO Model Under 70% TRAP.***

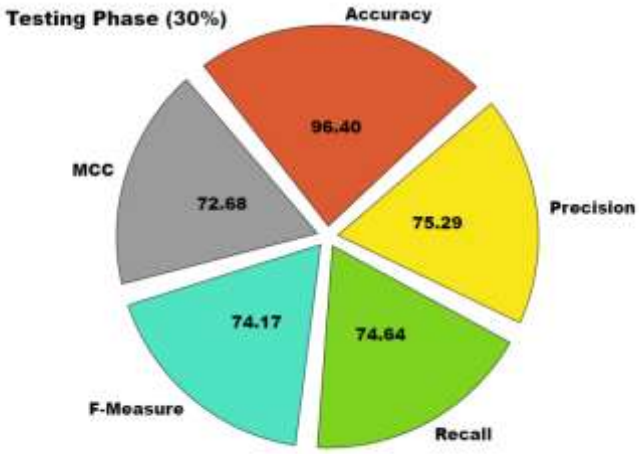


*Figure 5: Average Outcome of WCBGARNN-ICO Technique Under 30% TESP.*

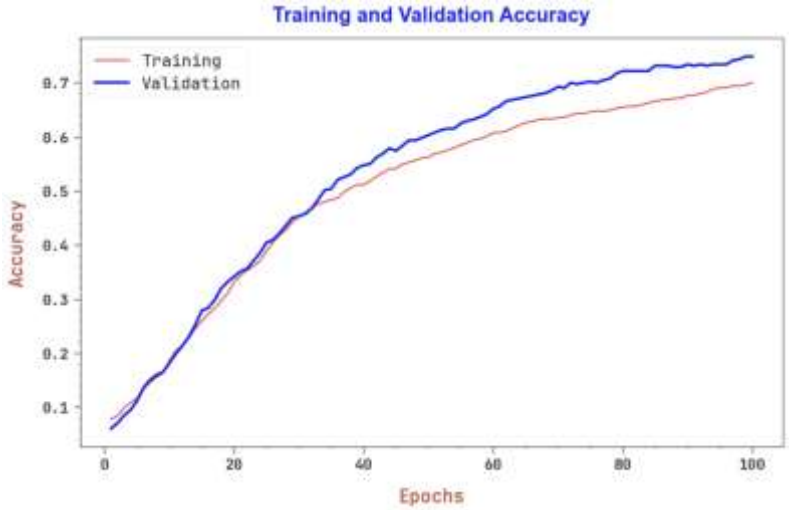


*Figure 6: Accu_Y Curve of WCBGARNN-ICO Technique.*

In Fig. 6, the training accu_y(TRAAC) and validation accu_y(VLAAC) values of the WCBGARNN-ICO system are exhibited. The rate of accu_y are estimated for 0-100 epoch counts. The figure emphasized that the values of TRAAC and VLAAC display an increasing trend that reported the capability of the WCBGARNN-ICO technique with superior execution over various iterations. Furthermore, the TRAAC and VLAAC stay nearer over the epochs, which shows lower minimum overfitting and demonstrates greater execution of the WCBGARNN-ICO algorithm, promising constant prediction on hidden samples.

In Fig. 7, the TRA loss (TRALS) and VLA loss (VLALS) graphs of the WCBGARNN-ICO system can be presented. The rate of loss is estimated for 0-100 epoch counts. It is denoted that the values of TRALS and VLALS show a lower trend, reporting the ability of the WCBGARNN-ICO methodology to balance a trade-off among generalization and data fitting. The incessant reduction in rate of loss furthermore promises the greater execution of the WCBGARNN-ICO method and fine-tuning the prediction values over time.

Table 4 and Fig. 8 exhibit the comparative findings of the WCBGARNN-ICO model with current approaches [23]. The performance value indicated that the WCBGARNN-ICO technique outperformed superior execution with several measures like ccu_y, prec_n, reca_l, and F_measure. In terms of accu_y, the WCBGARNN-ICO system got greater accu_y of 96.40% whereas the existing models such as NB, SVM, LSTM, BiLSTM, CE-WPC BERT, GloVe-Stacked BiLSTM, and BERT-SoftMax systems have got lower accu_y of 79.00%, 91.33%, 77.57%, 90.07%, 82.07%, 85.32%, and 95.63%, correspondingly. Also, for prec_n, the WCBGARNN-ICO model has greater prec_n of 75.29% whereas NB, SVM, LSTM, BiLSTM, CE-WPC BERT, GloVe-Stacked BiLSTM, and BERT-SoftMax methods have the least prec_n of 65.97%,

72.17%, 70.29%, 66.05%, 69.41%, 65.54%, and 74.48%, respectively.

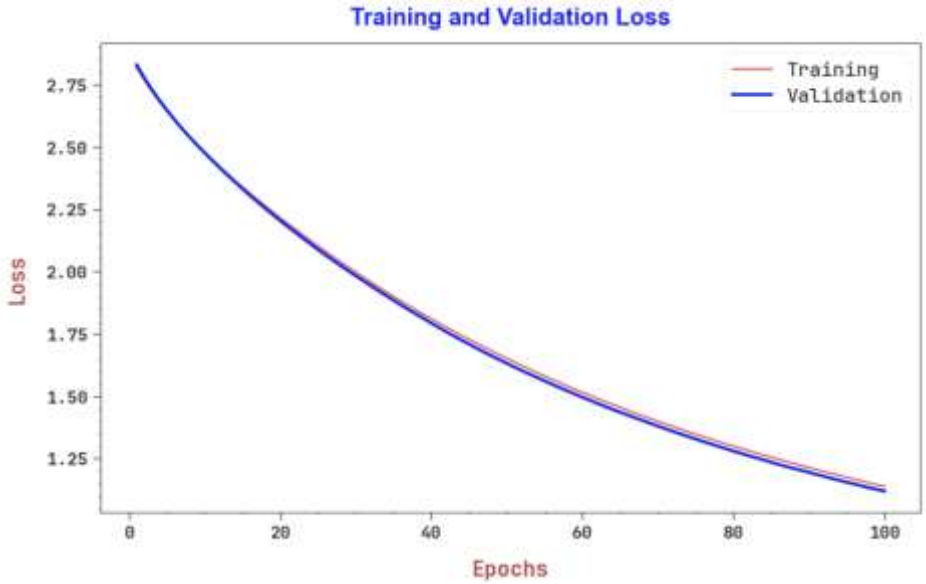


*Figure 7: Loss Curve of WCBGARNN-ICO Technique.*

*Table 4: Comparative Study of WCBGARNN-ICO Model with Existing Approaches.*

| Models | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{measure}$ |
|---|---|---|---|---|
| Naive Bayes Model | 79.00 | 65.97 | 67.04 | 67.15 |
| Support Vector Machine | 91.33 | 72.17 | 72.02 | 72.51 |
| LSTM Classifier | 77.57 | 70.29 | 73.50 | 72.53 |
| BiLSTM Model | 90.07 | 66.05 | 73.67 | 67.55 |
| CE-WPC BERT | 82.07 | 69.41 | 72.19 | 73.69 |
| GloVe-Stacked BiLSTM | 85.32 | 65.54 | 68.47 | 71.38 |
| BERT-SoftMax | 95.63 | 74.48 | 73.36 | 73.73 |
| WCBGARNN-ICO | 96.40 | 75.29 | 74.64 | 74.17 |

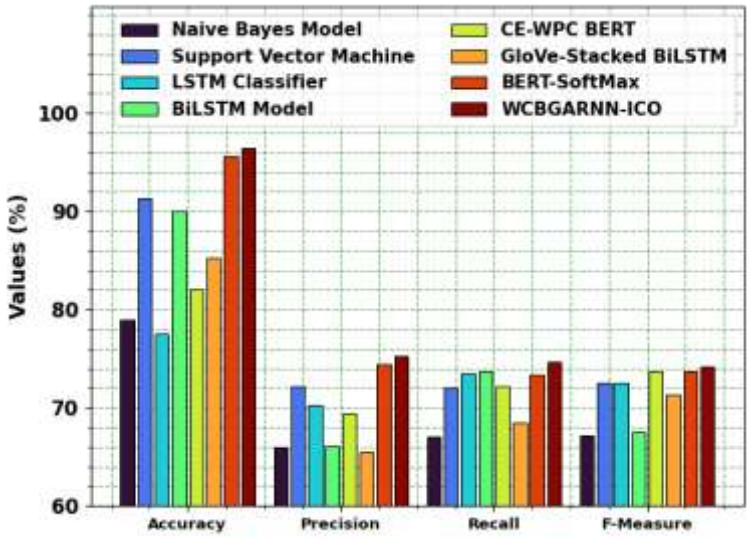


*Figure 8: Comparative Analysis of WCBGARNN-ICO Approach with Existing Approaches.*

In Table 5 and Fig. 9, the comparative study of the WCBGARNN-ICO algorithm is indicated based on implementation time (IT). The findings stated that the WCBGARNN-ICO methodology got superior outcomes. In terms of IT, the WCBGARNN-ICO system offers lowest IT of 6.14min while the NB, SVM, LSTM, BiLSTM, CE-WPC BERT, GloVe-Stacked BiLSTM, and BERT-SoftMax models get

superior IT values of 18.16min, 12.53min, 15.51min, 17.07min, 17.75min, 13.43min, and 7.47min, respectively.

*Table 5: IT Outcome of WCBGARNN-ICO Technique with Recent Models.*

| Models | Implementation Time (min) |
|---|---|
| Naive Bayes Model | 18.16 |
| Support Vector Machine | 12.53 |
| LSTM Classifier | 15.51 |
| BiLSTM Model | 17.07 |
| CE-WPC BERT | 17.75 |
| GloVe-Stacked BiLSTM | 13.43 |
| BERT-SoftMax | 7.47 |
| WCBGARNN-ICO | 6.14 |

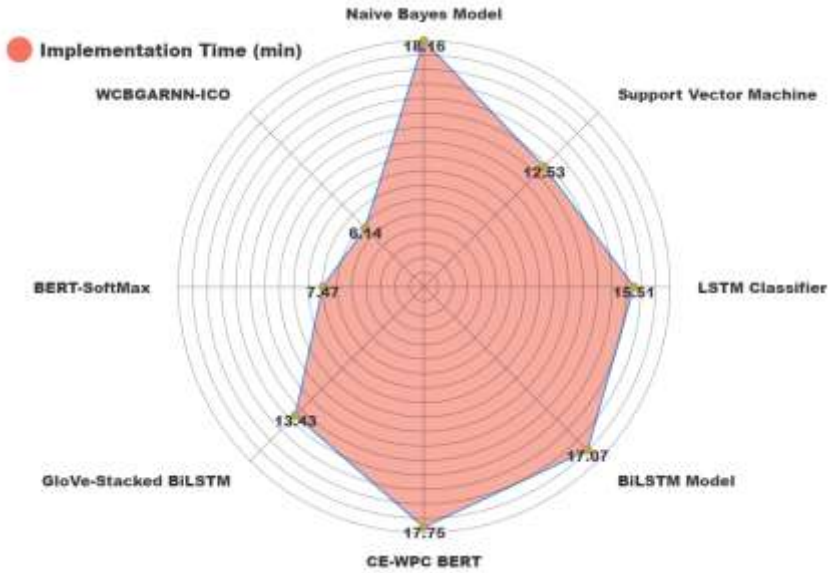


*Figure 9: IT Outcome of WCBGARNN-ICO Technique with Recent Models.*

## 5. CONCLUSION

In this manuscript, we have presented anew WCBGARNN-ICO technique. The key intention of the WCBGARNN-ICO model is to classify website in web mining. To achieve that, the WCBGARNN-ICO system has text preprocessing, BERT feature extractor, A-BGRNN based classification, and ICOA based parameter tuning processes. At first, the WCBGARNN-ICO model applies the quality of text preprocessing with different levels to attain clear data and extract significant data. For the extraction of feature process, the BERT method can be exploited. Besides, the A-BGRNN model is employed for the website classification process. At last, the ICOA is deployed to fine-tune the hyperparameter of A-BGRNN technique. The experimental study of the WCBGARNN-ICO algorithm can be tested on benchmark database and the outcome are measured with various measures. The experimental findings highlighted the development of the WCBGARNN-ICO system over other approaches.

**Funding Declaration:** No funding required.

## REFERENCES

[1] Espinosa-Leal, L., Akusok, A., Lendasse, A. and Björk, K.M., 2021. Website classification from webpage renders. In *Proceedings of ELM2019 9* (pp. 41-50). Springer International Publishing.

[2] Gupta, A. and Bhatia, R., 2021. Ensemble approach for web page classification. *Multimedia Tools and Applications*, 80(16), pp.25219-25240.

[3] Buncher, B. and Carrasco Kind, M., 2020. Probabilistic cosmic web classification using fast-generated training data. *Monthly Notices of the Royal Astronomical Society*, 497(4), pp.5041-5060.

[4] Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A. and Chen, A., 2020, April. Teachable machine: Approachable Web-based tool for exploring machine learning classification. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1-8).

[5] Li, Y.; Yang, Z.; Chen, X.; Yuan, H.; Liu, W., 2019. A Stacking Model Using URL and HTML Features for Phishing Webpage Detection. *Future Gener. Comput. Syst.*, 94, 27–39.

[6] Saleem Raja, A.; Vinodini, R.; Kavitha, A., 2021. Lexical Features Based Malicious URL Detection Using Machine Learning Techniques. *Mater. Today Proc.*, 47, 163–166.

[7] Shabudin, S., Sani, N.S., Ariffin, K.A.Z. and Aliff, M., 2020. Feature selection for phishing website classification. *International Journal of Advanced Computer Science and Applications*, 11(4).

[8] Murty, C.A. and Rughani, P.H., 2022. Dark web text classification by learning through SVM optimization. *J AdvInfTechnol*, 13(6).

[9] Demirkıran, F., Çayır, A., Ünal, U. and Dağ, H., 2020, September. Website category classification using fine-tuned BERT language model. In *2020 5th International Conference on Computer Science and Engineering (UBMK)* (pp. 333-336). IEEE.

[10] Zhang, X., Liu, J., Shi, M. and Cao, B., 2021, September. Word embedding-based web service representations for classification and clustering. In *2021 IEEE International Conference on Services Computing (SCC)* (pp. 34-43). IEEE.

[11] Tavasoli, M., Lee, E., Mousavi, Y., Pasandi, H.B. and Fekih, A., 2024. Wipe: A novel web-based intelligent packaging evaluation via machine learning and association mining. *IEEE Access*.

[12] Rao, A.S. and Venkatraman, P., 2024. Open Sources Tools in Web Mining Method. *Journal of Computer Graphics and Multimedia Applications*, p.10.

[13] Afif, M., 2023. Exploring the quality of the higher educational institution website using data mining techniques. *Decision Science Letters*, 12(2), pp.279-290.

[14] Amane, M., Aissaoui, K. and Berrada, M., 2023. Enhancing learning object analysis through fuzzy C-means clustering and web mining methods. *Emerging Science Journal*, 7(3), pp.799-807.

[15] Zhao, W., Liu, X., Xu, R., Xiao, L. and Li, M., 2024. E-commerce webpage recommendation scheme base on semantic mining and neural networks. *arXiv preprint arXiv:2409.07033*.

[16] Zaidi, T., Kumar, A. and Pundeer, S., 2024, May. Shifting From Syntactic to Semantics for Knowledge Exemplification Using Semantic Web (SW) Mining Techniques. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 167-171). IEEE.

[17] SheykhAbbasi, B., Abdolvand, N. and Rajaee Harandi, S., 2022. Predicting Customers' Behavior Using Web-Content Mining and Web-Usage Mining. *International Journal of Information Science and Management (IJISM)*, 20(3), pp.141-163.

[18] Rouxel, A., 2020. Media Content Analysis of Covid-19 Virus Using Natural Language Processing Techniques (Doctoral dissertation, Dublin, National College of Ireland).

[19] Zivkovic, M., Jovanovic, L., Bukumira, M., Antonijevic, M., Mladenovic, D. and Al, M., 2024, August. Optimizing SQL injection detection using BERT encoding and AdaBoost Classification. In *2nd International Conference on Innovation in Information Technology and Business (ICIITB 2024)* (pp. 137-154). Atlantis Press.

[20] Yu, Q., Zhao, H. and Wang, Z., 2019, August. Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. In *Proceedings of the 2nd international conference on artificial intelligence and pattern recognition* (pp. 116-119).

[21] Qi, Z., Yingjie, D., Shan, Y., Xu, L., Dongcheng, H. and Guoqi, X., 2024. An improved Coati Optimization Algorithm with multiple strategies for engineering design optimization problems. *Scientific Reports*, 14(1), p.20435.

[22] Mehta, H. (n.d.). Website Classification [Dataset]. Kaggle. https://www.kaggle.com/datasets/hetulmehta/website-classification

[23] Nandanwar, A.K. and Choudhary, J., 2023. Contextual embeddings-based web page categorization using the fine-tune BERT model. *Symmetry*, 15(2), p.395.

[24] Krosuri, L. R., & Aravapalli, R. S. (2024). Novel heuristic bidirectional-recurrent neural network framework for multiclass sentiment analysis classification using coot optimization. Multimedia Tools and Applications, 83(5), 13637-13657.

[25] Deivakani, M. (2025). Anomaly Detection in IoT Network Traffic Using Bidirectional 3D Quasi-Recurrent Neural Network Optimize With Coati Optimization Algorithm. Transactions on Emerging Telecommunications Technologies, 36(1), e70026.

[26] Dwivedi, S. N., Karnick, H. C., & Jain, R. (2026). A novel position attention-based convolutional bidirectional-gated network for text relation classification. Knowledge and Information Systems, 68(1), 18.