International Journal in IT & Engineering (IJITE)
Volume 13 Issue 02, February 2026 ISSN: 2321-1776 Impact Factor: 8.256
Journal Homepage: http://ijmr.net.in, Email: irjmss@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal

A BUTTERFLY-OPTIMIZED BbLSTM ALGORITHM FOR CLASSIFYING USER-VISITED WEB PAGES

Prayas Patel[1], Bharti Kumari[2], Mukesh Yadav[3]

[1]Research Scholar, Department of Computer science and engineering, SHEAT College of Engineering, Varanasi, UP

[2,3]Assistant Professor, Department of Computer science and engineering, SHEAT College of Engineering, Varanasi, UP

ABSTRACT

Overall, multi-class web page categorization based on webpage content utilizing LSTM deep learning architecture and butterfly optimizer is a promising strategy that outperforms current state-of-the-art web page classification models. To achieve this purpose, this paper offers a Butterfly-based Long Short-Term Memory Classifier (BbLSTMC) architecture for multi-class classification of web pages based on content, which may be used to a variety of domains such as e-commerce, news, social media, and so on. The BbLSTMC detected findings for user-visited web categories flow effortlessly into the next step of the proposed paradigm. Developers may exploit the Butterfly-based LSTM architecture's capabilities to create new solutions that will help organizations improve their web presence, improve the user experience, and boost consumer engagement.

Keywords: LSTM, Deep Learning Architecture, Butterfly, Optimizer

INTRODUCTION

Machine learning algorithms are vital for effectively categorizing and structuring web pages based on their content and information. They automate the examination and classification of massive amounts of online data, allowing for more efficient page aggregation. Nonetheless, the online pages include implicit and contextual information that significantly influences their categorization.

To collect and interpret these deep patterns and common qualities, machine learning algorithms must be constantly taught and refined. Furthermore, developments in machine learning have enabled the introduction of deep learning algorithms, which have showed outstanding skill in handling complicated online data and extracting intricate patterns and high-level characteristics.

Deep learning-based web site classification provides various advantages over machine learning techniques, including the capacity to learn complicated representations, better handling of sequential data, more resilience to noise and fluctuations, automated feature extraction, and improved accuracy. RNNs have demonstrated promising results in web page categorization challenges due to their capacity to interpret sequential input. RNNs are designed to handle time-varying data such as text, voice, and video. RNNs are effective in classifying web pages because they can use sequential data to identify long-term relationships. RNNs are well-suited to this task since web site content is frequently structured hierarchically.

LSTM cells are a form of RNN designed to address the issues of disappearing and exploding gradients that arise while training deep neural networks. LSTM cells can select whether to

preserve or discard information based on the current input and their previous state. This makes them ideal for modelling long-term dependencies in sequential data. The training optimization procedure, however, has a significant impact on the LSTM architecture's perform

Optimization methods are essential for training deep neural networks, including LSTM. Traditional optimization techniques are limited in terms of completion speed as well as efficiency. This condition has resulted in the creation of novel optimization techniques that can increase the efficacy of training and accuracy of LSTM architectures.

To address these issues, a new optimization technique known as the Butterfly Optimizer has been introduced to speed up and increase the accuracy of deep neural network training. The Butterfly Optimizer adjusts the learning rate during training by moving like a butterfly. This results in faster convergence and higher accuracy compared to conventional optimization techniques.

The ML-CWUBS uses a unique technique to classify user-visited web pages. It employs a Long Short Term Memory (LSTM) model alongside a Butterfly Optimizer for Classification (BbLSTMC). This model captures the local and global context of web page content by integrating RNNs' sequence-based characteristics with LSTM memory retention capabilities.

The Butterfly Optimizer additionally adjusts the model's weights and biases to boost performance. On a real-world web page dataset, the suggested technique outperforms even the most powerful algorithms in terms of accuracy, precision, and recall. The BbLSTMC approach shows promise in accurately and quickly categorizing web pages with multiple classes.

REVIEW OF RELATED LITERATURE

F. Mostajabi et al. [1] explored the shift from data mining to big data and identified the underlying difficulty of analyzing enormous volumes of data to extract useful information or knowledge for future actions. Furthermore, they examined current big data technologies and attempted to assist in determining the appropriate mix of these technologies depending on particular application requirements. The researchers also noted the extensive use of big data in several businesses.

Y. Roh et al. [2] investigated machine learning and data management integration in the context of big data. These publications discussed the most current advancements in machine learning for large data processing, as well as the difficulties they raised. The cause-and-effect relationship between these challenges was studied by classifying them based on big data dimensions such as volume, velocity, diversity, and veracity. Furthermore, the study emphasized the development of machine learning methodologies and techniques that may help practitioners overcome these challenges and pick appropriate solutions for their specific use cases. Furthermore, resilient computing methodologies were investigated, and future research opportunities as prospective areas of investigation in this field were identified.

A. Duque Barrachina and A. O'Driscoll [3] suggested that technical assistance contact centres are another application field for big data analytics. Furthermore, the internet has become a significant resource as a result of the rapid increase of information sources on the WWW, which has led to the widespread use of online commerce.

Iacopo P. et al. [4] looked at the problem of distinguishing between human and robot conduct in internet networks. They proposed a unique method for assessing robot and human behaviour

based on network activity and statistical analyses. This method enables more precise identification of robot activity and better insights into the dynamics of human-robot interactions. The authors stressed the need of understanding robots' influence on social and economic systems and developing appropriate methods to mitigate their negative implications.

B. W. et al. [5] explored critical issues in web robot identification and offered unique ways for finding attributes that might distinguish web robots from human users in real time. They used machine learning techniques to show their empirical findings on real-world datasets. They discovered that more sophisticated algorithms are necessary to detect advanced online robot behaviours in a variety of applications, including network traffic monitoring, website security, social media analysis, marketing research, and consumer feedback analysis.

According to M. Gan and K. Xiao [6], combining RNN and LSTM models shows significant potential in web page categorization assignments. RNNs can detect sequential relationships in data, which is useful for studying the sequential nature of web page content. The LSTM model increases the capacity to describe long-term dependencies in web page content by including memory cells capable of selectively retaining and forgetting information over time.

L. Wen et al. [7] investigated the use of butterfly optimisation approaches in various forms, with a focus on the procedure for picking features from high-dimensional datasets. The aforementioned research projects have used adaptive parameters and advanced mutation techniques to improve the efficacy and rate of convergence of the traditional butterfly optimisation algorithm. The use of these variances has been studied in a variety of fields, including but not limited to renewable energy, cancer categorization, and gene expression analysis.

Y. Ren et al. [8] introduced pattern recognition algorithms to multilayer networks, revealing subtle interactions between connected layers. Pattern discovery has also contributed in the development of discovery systems, enabling autonomous task selection based on user behaviour patterns. Through the evaluation of historical data, these methodologies have contributed to a better understanding of train delays in the transportation domain. Furthermore, pattern discovery has played an essential part in improving company profitability through information discovery, as well as the analysis of patterns in internet background radiation.

W. M. Kouw and M. Loog [9] investigated domain adaptation and its relevance to pattern analysis in a machine learning context. The authors grouped numerous methodologies into three categories: sample-based, feature-based, and inference-based, all of which are necessary for pattern analysis. The authors provided major prospects for future research in the domain of domain adaptation by categorizing and clarifying their relationship with pattern analysis.

Akarsh et al. [10] studied malware family identification. In their respective domains, both research emphasized the need of interpretation and visualization methods. Using these strategies, researchers may get useful insights, enhance decision-making processes, and broaden their understanding of deep learning models. The outcomes of these research help to shape interpretation and visualization approaches, paving the way for more effective implementations in their respective fields.

Amiri, Z., et al. [11] investigate the rationales, techniques, and timing of human technology adoption across domains such as the Internet of Things (IoT), behavioural science, and edge

analytics. As a consequence of the lack of comprehensive research into the use of ML-based techniques in the context of using IoB for medical uses, we performed a study on the issue, providing a novel taxonomy that emphasizes the need of employing each ML method uniquely. With this goal in mind, we have divided the cutting-edge ML solutions for IoB-based healthcare issues into five groups. These include convolution neural networks (CNNs), recurrent neural network (RNNs), deep neural networks (DNNs), multilayer perception (MLPs), and hybrid approaches.

Rane, N. L., et al., [12] undertake a detailed analysis of ML methods suited for huge datasets, including approaches to reinforcement learning, unsupervised learning, and supervised learning. A research is also conducted on several deep learning structures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, which are recognized for their capacity to detect complicated patterns in datasets with an examining emerging advancements such as automated machine learning, edge computing, and the possible integration of quantum computing with ML sheds light on the future of sophisticated data analysis. To guarantee that these technologies are used responsibly, ethical considerations such as data privacy, bias, and model interpretability are rigorously examined and addressed. This publication intends to provide a comprehensive resource for scholars and practitioners interested in using machine learning and deep learning for sophisticated analyses of data.

Rani, S. et al. [13] discuss the relevance of big data analysis and machine learning in improving diagnosis, operational efficiency, and tailored patient care. It investigates the main difficulties of data heterogeneity, privacy, the complexity of computation, and advanced methodologies including federated learning (FL) and edge computing. Real-world applications, including as illness prediction, medical imaging, medication development, and remote surveillance, demonstrate how ML approaches like deep learning (DL) and natural language processing (NLP) improve clinical decision-making. The study stresses the practical implementation of intelligent systems, including case examples that show up to 95% diagnosis accuracy and cost savings.

Agrawal, D. et al. [14] introduce a unique Deep Reinforcement Learning (DRL) model for credible stock price forecasting using financial time series data, which is based on the Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) architectures. Backtesting shows larger cumulative returns, stronger Sharpe ratios, and lower maximum drawdowns, implying bigger profits and lower risk. Experimental results based on comparisons with the state-of-the-art demonstrate the suggested approach's resilience and generality, implying that it is promising for real-world automated trading and financial decision-making in dynamic markets.

Abduljabbar et al. [15] developed and evaluated BiLSTM short-term traffic forecasting models. To this end, they utilized data derived from a calibrated micro-simulation model developed for a congested freeway in Melbourne, Australia. Subsequently, loop detector data—comprising speed, flow, and occupancy—was generated using the base-year simulation model. This data was employed to develop and compare various LSTM models for forecasting short-term traffic up to 60 minutes into the future. The modeling results indicated that, under base-year conditions, the BiLSTM model outperformed other forecasting models across several distinct prediction horizons. Subsequently, the simulation model was adapted to reflect future-year scenarios; in these scenarios, traffic demand was increased by 25% to 100% to simulate potential future growth in traffic demand. The results demonstrated the superior performance of the BiLSTM model across all traffic variables and across multiple prediction horizons.

Wang, W., etal., [16] develop an Internet of Things (IoT)-enabled framework for sustainable product design that enhances eco-efficiency through transparent, data-driven decision-making. The BiLSTM model is benchmarked against LSTM, CNN, and traditional machine learning techniques to assess predictive performance. Robustness is ensured using five-fold cross-validation and statistical significance testing (t-test, $p < 0.05$). Results indicate that the proposed framework improves energy efficiency by 23.5% and reduces material waste by 19.2% compared to conventional methods. The approach is applicable across automotive, electronics, and consumer goods sectors and supports measurable progress toward the United Nations Sustainable Development Goals (SDGs).

Shi, P., et al. [17] proposes a dissolved oxygen prediction method based on an improved sparrow search algorithm optimized TCN- BiLSTM (SMI-TCN BiLSTM). Initially, the Savitzky-Golay (SG) filter is employed to denoise the water quality data, producing smoother and more consistent datasets. In addition, the traditional Temporal Convolutional Network (TCN) often fails to capture the dynamic fluctuations present in DO data, resulting in suboptimal prediction performance. To overcome this limitation, a Bi-directional Long Short-Term Memory (BiLSTM) network is integrated into the TCN framework, forming a TCN-BiLSTM prediction module. This module effectively captures both forward and backward temporal dependencies, improving the model's ability to track the dynamic trends in the data and enhancing its prediction accuracy.

Deepak H a,etal., [18] presents an automated web scraping technique that blends a feature selection model with deep learning. The pre-processed data is then utilised to train a text-based Convolutional Neural Network (CNN) model, which is then used to extract features. The characteristics are then picked in the most optimal manner using the Butterfly Optimisation Algorithm (BOA) model. way. The chosen traits are then categorised as Long-Short Term Memory (LSTM). In our studies, the suggested model is assessed using the WebKB benchmark dataset. Our experimental find-ings demonstrated that the suggested model was frequently able to learn the appropriate decision surface using the area under the Precision-Recall Curve as the performance met-ric.

OBJECTIVES

The main aim of this study is to develop a Butterfly-Optimized BbLSTM Algorithm for the Classification of Web Pages Visited by Users

OPTIMAL CLASSIFICATION MODEL USING RNN ARCHITECTURES

Recently, the deep learning age has witnessed incredible advancements, with RNNs emerging as a powerful tool for modelling sequential data. RNNs, which are differentiated by their capacity to record temporal correlations, have shown significant promise in a wide range of applications, most notably classification jobs. The advantages of utilizing RNNs for optimal classification models are many and powerful. The primary goals are to capture long-term dependencies, deal with variable-length input sequences, learn from contextual information, make sequential decisions, and be successful for multi-class categorization.

Exploring these motivations reveals that RNNs have many distinct benefits that can significantly improve classification model performance. RNNs, in particular, provide distinct advantages and play an important role in improving the performance of web page categorization algorithms.

ARCHITECTURE OF PROPOSED BbLSTMC

After preprocessing, the proposed model focuses on grouping user-visited web pages based on their content. Deep learning methods, particularly recurrent neural networks, are appropriate for this task since web page content contains sequential data. Despite their superior learning capabilities, deep learning RNNs have setup and training hurdles due to disappearing and ballooning gradient concerns. Deep RNN training is based on extracting higher-level nonlinear characteristics from a large amount of sequential data and treating it as a nonconvex optimization problem. This task is designed to minimize nonlinear loss functions with several local optimums.

To address this issue, the current study proposes a "Butterfly-based LSTM architecture to classify user-visited web page classification" (figure 1), which combines the benefits of (i) LSTM for developing an optimal RNN architecture to gain a competitive advantage in short-term prediction while managing better long-term prediction (ii) A novel meta-heuristic Butterfly Optimizer to handle gradient issues with guaranteed convergence to the global optima. (iii) Multi-class classification is a simpler technique to web page categorization that assigns a single label or category to each page based on the most relevant or prominent subject.

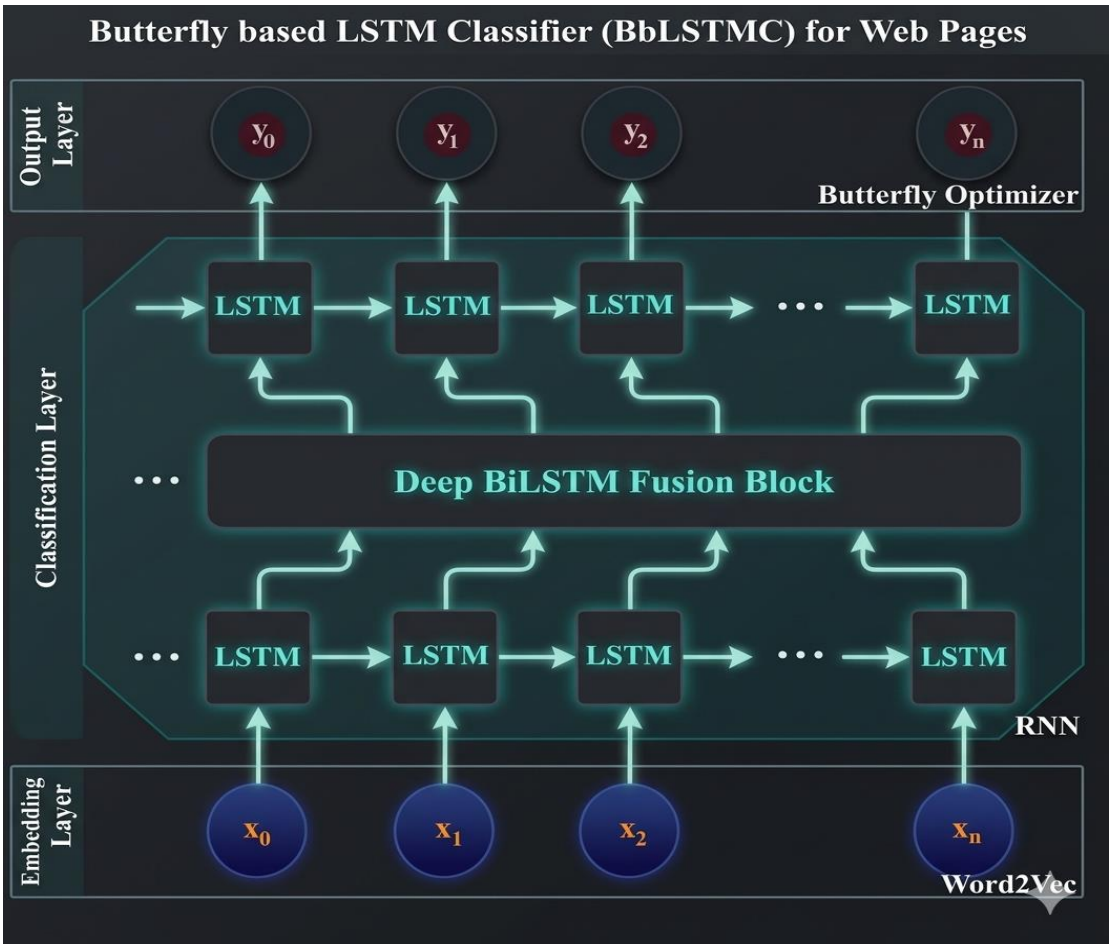

FIG. 1: ARCHITECTURE OF PROPOSED BbLSTMC

The process of retrieving specific pieces of information or material from a website is known as data extraction from viewed web pages. The specific requirements and obstacles of the aim, such as the type of material, the frequency of changes, moral and legal concerns, technological competence, and scalability, all influence which approach is the most successful for extracting content from visited web sites. There are a few common ways for obtaining material from previously viewed internet pages. Web scraping is one of the techniques provided in the suggested paradigm.

Web scraping refers to the technique of obtaining specified data from websites using automated technology. One approach to achieving this aim is to employ software tools that can "parse" a web page's HTML code and extract usable data depending on specified criteria. The suggested model parses the page's HTML content using the BeautifulSoup Python module. The BbLSTMC method enables the suggested model to extract webpage content depending on the user's visited URL.

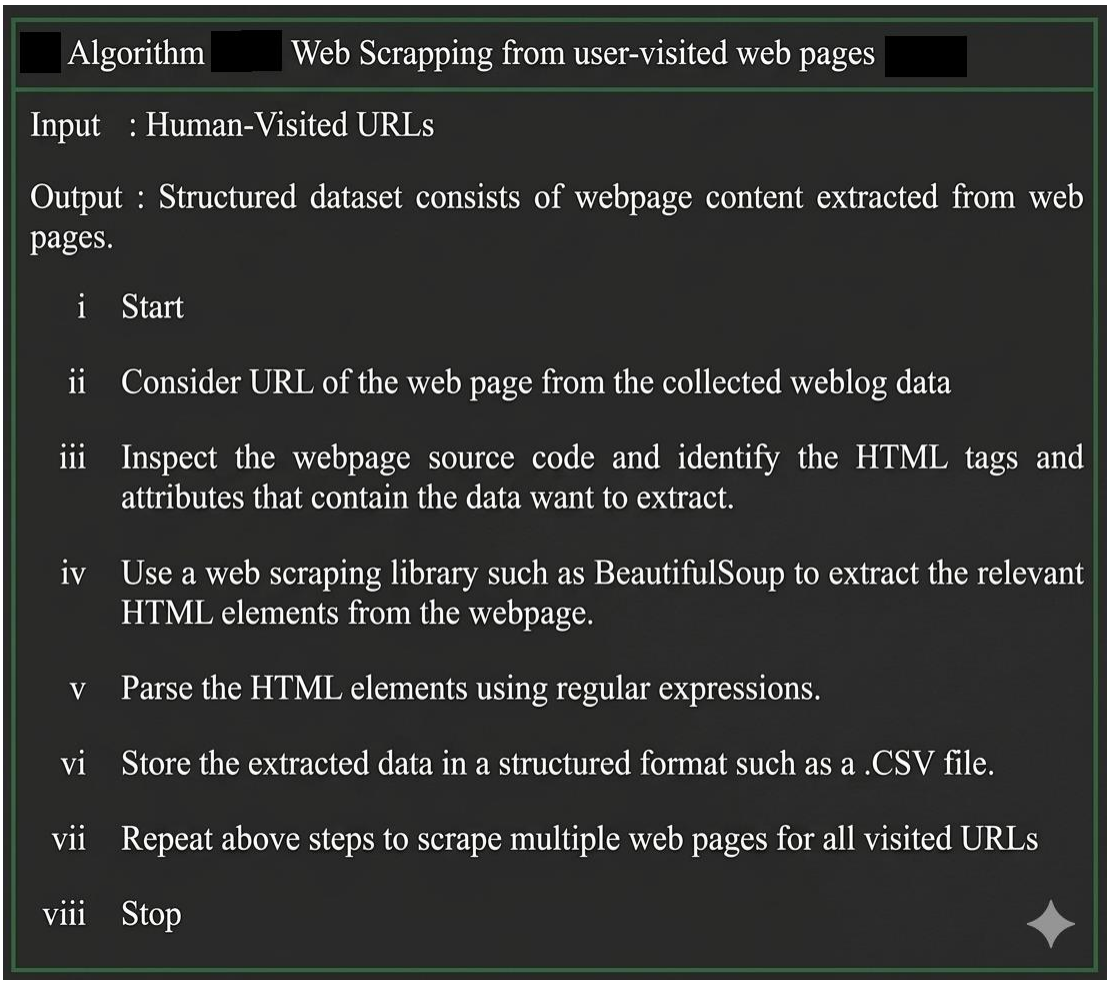

FIG. 2: ALGORITHM OF BbLSTMC

The algorithm-outlined technique is critical to the proposed research, serving as a dependable and helpful instrument for systematically analyzing and collecting text from the meta tags of web

pages accessed by humans via specific URLs. This initial step is an essential part of the proposed model's data refinement pipeline. These meta elements frequently include critical metadata, including as titles, descriptions, keywords, and other important information about the website's content.

This information is required to gain a full knowledge of the context and content of the pages being considered within the scope of the proposed model. The retrieved data is sent into BbLSTMC's embedding layer, which converts the text to vectors and prepares it for use with deep learning techniques.

Word embedding is a key NLP and deep learning approach. It requires transforming vocabulary words or phrases into high-dimensional vectors, usually in a continuous space. Word embedding methods such as Word2Vec, GloVe, and FastText aim to capture semantic links between words.

Because these vector representations incorporate information about word similarity and context, they allow algorithms to interact with words in more meaningful ways. Word embeddings have transformed NLP jobs by allowing models to understand the underlying meaning of words in a text, resulting in substantial advances in applications like sentiment analysis, machine translation, and text generation.

Word2Vec is a promising two-layer neural network for word vectorization when compared to other word embedding approaches. Word2vec provides distributed numerical representations of word characteristics, including context for individual words. If a word's assigned feature vector cannot properly predict its context, its portions are modified depending on error signals received from each word's context in the corpus. The numbers in the vectors of words with similar contexts are then modified automatically to bring them closer together. Word2vec is not a deep neural network, but it does convert text to numbers that deep neural networks can understand.

This approach was able to encode web page text as fixed-length vectors for LSTM using simple data processing methods such as stop word removal, stemming, and number sequence padding. This technology turns words into vectors using a linear combination of weighted averages, requiring little computer resources and ensuring great dependability. Cosine similarity is the best mathematical answer for determining the similarity of words scattered in geographic places, and it ranges from 0 to 1.

In the current model, Word2Vec first trains on a large corpus to obtain a quality distributed representation of each term on the web page. Equation 1 shows how the web page keyword vector is represented.

$$PV = \sum_{i=1}^{n} Word2Vec(kw_i) \qquad (1)$$

where $Word2Vec$ (kwi) is the ith term in the vector, and n is the total number of keywords on the web page. After successfully processing the embedding layer, each web page is represented as a vector, consisting of keywords PV = $\{kw1, kw2,..., kwn\}$. The resulting characteristics are sent to the classification layer, which then classifies the web pages.

BUTTERFLY BASED LSTM AT CLASSIFICATION LAYER

Machine learning offers a variety of models to assist with categorization problems, including bagging and boosting. In general, classification jobs are within ML's scope. However, when there are several output classes and a substantial quantity of data is necessary to support the model's performance, neural networks outperform. Common neural networks nowadays focus entirely on the current input data and forget all they've learnt in the past. Deep learning, on the other hand, is based on neural networks that have several layers of coupled neurons. Data is processed nonlinearly at these layers, allowing the network to learn and detect previously unknown patterns.

RNNs are deep neural networks that are very adept at modelling sequence data. However, when it comes to classification tasks, typical RNNs with short-term memory might have problems such as gradients ballooning or disappearing. To overcome these issues, the RNN model can use LSTM, which can efficiently detect long-distance correlations between terms on websites. Prominent researchers and industry professionals have attested to the advantages of integrating RNNs with LSTMs, which can improve the learning capacities of a variety of applications.

Each LSTM unit includes a memory cell and a state at time "t". LSTM gates regulate reading, writing, and deletion. It allows you to input, forget, and output using its expanded memory. At time 't', the LSTM unit gets two sources: your previous hidden state (ht-1), and the input vector (xt). In LSTM, each gate has an internal source, specifically the cell state Ct−1 of its Cell Block. Peephole communications are made between cells and their own gates.

The hidden layer state (ht) is determined by considering the internal cells of the gates and peephole connections, as well as the node status (xt, ht−1).

Thus, the hidden layer of classic RNN has been combined with LSTM as it matures into a more powerful framework for the current study challenge.

The hidden states of the the model suggested can recover linked behaviours and long-distance patterns via cyclic connections between neuron in each layer (keywords on the web page).

Given that web page vector $x_t = \{kw_1, kw_2, \ldots, kw_n\}$ are put into recurrent layer step-by-step. Each layer of LSTM is updated and summarized as follows:

The Input Gate $it = \sigma(W_{ix}\, x_t + W_{reci}\, h_{t-1} + W_{ic}\, c_{t-1} + b_i)$ (2)

The Forgotten Gate $f_t = \sigma(W_{fx}\, x_t + W_{recf} h_{t-1} + W_{fc}\, c_{t-1} + b_f)$ (3)

The Cell State $c_t = f(t) \odot c_{t-1} + i_t \odot g(W_{cx}\, x_t + W_{recc} h_{t-1} + b_c)$ (4)

The Output Gate $O_t = \sigma(W_{ox}\, x_t + W_{reco} h_{t-1} + W_{oc}\, c_t + b_o)$ (5)

The Hidden State $h_t = o_t \odot m\,(c_t)$ (6)

In the above equations,

W_{ic} , W_{fc} , W_{oc} are peephole connections

W_{ix} , W_{fx} ,W_{cx} , W_{ox} are input weight matrix connections

W_{reci} ,W_{recf} ,W_{recc} , W_{reco} are the recurrent weight matrix connections

b_i , b_f , b_c , b_o are the bias values

⊙ denotes element-wise multiplication

Even if RNN-LSTM is included in the proposed model, deep RNN learning methods are still impacted by starting weights and noise fluctuations throughout the dataset's nonconvex optimization. As a result, training a deep RNN model is dependent not just on long-term knowledge retention (via LSTM), but also on parameter beginning values. Most initialization methods employ random initialization, weights near to the identity matrix, or the conventional convention.

As a result, choosing the best initial parameters for a particular model and recognizing which parameters need to be modified remains a difficult optimization problem. This is mostly owing to a lack of exact understanding regarding which aspects of these parameters are preserved or learnt under different settings. As a result, this is still an open and tough optimization task in the area. To address the circumstances described above, the proposed work uses Butterfly Optimization to train a Deep RNN-based LSTM model.

DESIGN AND DEVELOPMENT OF BbLSTMC ALGORITHM

To solve gradient difficulties and avoid trapping of local optima in the classification layer, the optimizer must update the model in response to the loss function's output. There are nature-inspired algorithms that are used to solve global optimization issues, but they have complex limitations.

Butterfly optimization methodology is a meta heuristic method ideal for nonlinear challenges inspired by nature that mimics the food-seeking behaviour of butterflies, which has piqued the interest of many scholars.

According to biological knowledge and observations, a butterfly has the ability to emit its own particular aroma at a specific strength, which then travels across the region and is detected by others. Determining each butterfly's aroma requires an understanding of how sensory modalities such as smell are processed by stimulus. There are three main words used to define smell:

Sensory modality (c) - Sensory refers to assessing the sort of energy, whereas Modalities refers to the raw input used by the senses, in this case scent.

Stimulus Concentration (I) - The magnitude of a real stimulus related to the solution's fitness.

Power Exponent (a) – The exponent to which intensity is elevated is power.

$$Perceived\ Magnitude\ of\ the\ Smell\ (f) = c * I^a \qquad (7)$$

The process of the butterfly affects the optimizer's fitness. A butterfly will approach other butterfly with more endurance or a stronger fragrance, and the procedure will keep going until the global ideal is attained.

$$\textit{Global Optima} \qquad x_i^{t+1} = x_i^t + (r^2 * g^* - x_i^t) * \qquad (8)$$

When a butterfly cannot identify a scent stronger than its own, it will travel at random, which is comparable to local optima.

$$\textit{Local Optima} \qquad x_i^{t+1} = x_i^t + (r^2 * x_j^t - x_k^t) * f_i \qquad (9)$$

x_j^t and x_k^t are the solution vectors for j^{th}, k^{th} butterfly randomly chosen from solution space, where r is random number generated by the equation $r = rand(0,1)$

The classification layer improves the Butterfly optimizer's ability to properly categorize visited webpages. The butterfly's fitness determines the metadata of a webpage by detecting its aroma. The suggested model uses the following equation to classify web pages in the defined web category $WPCm$:

$$WPC_m = WL_{dt} \rightarrow \sum_{m=1}^{N} L_m Fe_m \left(B_f \left(O_w (m)\right)\right) \qquad (10)$$

Where,

WL_{dt} is weblog data set

L_m is the N^{th} quantity of web user visited pages

Fe_m is set of selected Features

B_f is Fitness of Butterfly Optimizer

O_w is classification of Web page category

The unidirectional nature of RNN-LSTM, along with the nature-inspired Butterfly optimizer, allows it to effortlessly balance the imbalanced amount of information stored in the hidden states, addressing classic RNN's bursting and disappearing gradients.

Finally, the representation of keywords of web pages is obtained by combining forward hidden state $\overrightarrow{h_t}$ and backward hidden state $\overleftarrow{h_t}$

i.e., $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$ (11)

The web page vector is represented as h = {h1, h2, … hn } before being sent to the output layer. The dimensions of 'h' are then modified to fit the output layer. To classify a given website into

one of numerous specified categories, RNN-LSTM ends in a penultimate layer that produces actual valued scores that are not easily scaled and may be difficult to cope with multi-class web page classification.

The output layer of the BbLSTMC employs the softmax activation approach to convert LSTM data to a probability distribution across categories. When an input is small or negative, the softmax assumes a small probability between 0 and 1. This is because the suggested model may utilize the network's output to compute the chance that a particular website would fall into each category. As a result, the proposed model includes the Softmax function as a final layer for multi-class classification of neural networks.

$$y_t = \sigma(h_t(i)) = \frac{e^{h_i}}{\sum_{j=1}^{k} e^{h_j}} \qquad (12)$$

The suggested model for web page categorization with LSTM networks employed the categorical cross-entropy loss technique as the cost function. This cost function is appropriate for multi-class classification issues in which each input is assigned to one of multiple classes.

The categorical cross-entropy is calculated as the difference between the projected probability distribution across classes and the input's real probability distribution (typically provided in one-hot encoded format). The network's purpose during training is to decrease this dissimilarity. The cost function is the categorical entropy loss function.

$$J(\theta) = -\frac{1}{m} \sum_{t=1}^{m} \bar{y_i} \log(y_t) + \lambda ||\theta||2^F \qquad (13)$$

Where $\bar{yi}$ is the ground truth of keyword, yi is the estimated probability for each class, m is the size of the dataset and λ is Regularization hyper parameter. According to the detailed analysis presented above, the algorithm 2 of BbLSTMC is as follows:

Algorithm 2: Butterfly based LSTM architecture for human-visited web page classification

Step 1. Input : Uncategorized web pages, initial parameters of BbLSTMC

Step 2. Output : Web pages into defined categories

Step 3. Initialize the LSTM model using starting parameters such as the number of LSTM cells, learning rate, and so on, as well as Butterfly Optimization Algorithm (BOA) parameters such as the population's size, maximum number of iterations.

Step 4. Pre-process the web page data by translating the text to numerical representations with Word2Vec, then divide the data between training and testing sets.

Step 5. Train the Deep RNN-LSTM model on the training set using the backpropagation algorithm with the initial parameters.

Step 6. Evaluate the efficacy of the BbLSTM classifier on the testing set to obtain an initial fitness score.

Step 7. Optimize the Deep RNNLSTM model's measurements using the BOA method. The BOA method works by imitating the behaviour of a bunch of butterflies as they search for the best solution. The method adjusts the RNN-LSTM model's parameters depending on the solutions' fitness

Step 8. Update the BbLSTMC model's parameters using the BOA method (from Eq (4.11) to Eq (4.14)) and retrain the model on a training set with the new variables.

Step 9. Use the Softmax activation function to transform LSTM data into a probability distribution across categories.

Step 10. Use the categorical cross-entropy loss function to measure the difference between the projected probability distributions.

Step 11. To determine a new fitness score, evaluate the modified RNN-LSTM the model's eff on the testing set.

Step 12. Repeat steps 5–9 until the fitness score of the RNN-LSTM model converges or the maximum number of iterations is reached.

Step 13. Use the optimized RNN-LSTM model in a web application to categorize web pages in real time.

CONCLUSION

The Butterfly-based RNN-LSTM architecture has demonstrated outstanding performance, highlighting the model's promise for web page multi-class categorization. The suggested model's amazing capabilities enable it to efficiently and precisely assess the properties of web pages within a reasonable timeframe. Its capacity to rapidly evaluate and filter through vast volumes of data makes it an excellent choice for a variety of applications. The BbLSTMC detected findings for user-visited web categories flow effortlessly into the next step of the proposed paradigm. Developers may exploit the Butterfly-based LSTM architecture's capabilities to create new solutions that will help organizations improve their web presence, improve the user experience, and boost consumer engagement.

REFERENCES

1. F. Mostajabi, A. A. Safaei, and A. Sahafi, (2021) "A Systematic Review of Data Models for the Big Data Problem," in IEEE Access, vol. 9, pp. 128889-128904, 2021.
2. Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 4, pp. 1328-1347, 2019.
3. A. Duque Barrachina and A. O'Driscoll (2014), "A big data methodology for categorizing technical support requests using Hadoop and Mahout," in Journal of Big Data, vol. 1, no. 1, pp. 1-11, 2014.

4. B. W. On, J. Y. Jo, H. Shin, J. Gim, G. S. Choi, S. M. Jung, (2021) "Efficient Sentiment-Aware Web Crawling Methods for Constructing Sentiment Dictionary," IEEE Access, vol. 9, pp. 161208-161223, 2021, DOI: 10.1109/ACCESS.2021.3129187.
5. M. Gan and K. Xiao, "R-RNN: Extracting User Recent Behavior Sequence for Click-Through Rate Prediction," IEEE Access, vol. 7, pp. 111767-111777, 2019, DOI: 10.1109/ACCESS.2019.292771
6. L. Wen, T. Wenbin, X. Ming, T. Mingzhu, and C. Shaohong, "Parameters identification of photovoltaic models by using an enhanced adaptive butterfly optimization algorithm," Energy, vol. 229, p. 120750, 2021, DOI : 10.1016/j.energy.2021.120750.
7. Y. Ren, A. Sarkar, P. Veltri, A. Ay, A. Dobra, and T. Kahveci, "Pattern discovery in multilayer networks," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 8, 2018, DOI: 10.1109/TCBB.2021.3105001.
8. W. M. Kouw and M. Loog (2019), "A Review of Domain Adaptation without Target Labels," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, pp. 766-785, DOI: 10.1109/TPAMI.2019.2945942.
9. Akarsh, S., Poornachandran, P., Menon, V. K., & Soman, K. P. (2019). A detailed investigation and analysis of deep learning architectures and visualization techniques for malware family identification. Cybersecurity and Secure Information Systems: Challenges and Solutions in Smart Environments, 241-286.
10. Amiri, Z., Heidari, A., Darbandi, M., Yazdani, Y., Jafari Navimipour, N., Esmaeilpour, M., ... & Unal, M. (2023). The personal health applications of machine learning techniques in the internet of behaviors. *Sustainability*, *15*(16), 12406.
11. Rane, N. L., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). Machine learning and deep learning for big data analytics: A review of methods and applications. *Partners Universal International Innovation Journal*, *2*(3), 172-197.
12. Rani, S., Kumar, R., Panda, B. S., Kumar, R., Muften, N. F., Abass, M. A., & Lozanović, J. (2025). Machine learning-powered smart healthcare systems in the era of big data: Applications, diagnostic insights, challenges, and ethical implications. *Diagnostics*, *15*(15), 1914.
13. Agrawal, D., Raja, R., Dewangan, D., Kambhampati, K., & Sahu, D. K. (2026). Deep reinforcement learning approach for reliable stock price prediction using financial time series data. In *Computational Techniques and Smart Manufacturing* (pp. 587-594). CRC Press.
14. Abduljabbar, R.L., Dia, H. & Tsai, PW. (2021) Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data. Sci Rep 11, 23899 (2021). https://doi.org/10.1038/s41598-021-03282-z
15. Wang, W., Wu, J., & Zhao, L. (2025). An IoT-enabled AI framework for sustainable product design optimizing eco-efficiency using BiLSTM. Scientific Reports.
16. Shi, P., Tang, M., Wang, Q. et al. (2025) Optimization of TCN-BiLSTM for dissolved oxygen prediction based on improved sparrow search algorithm. Sci Rep 15, 30790 (2025). https://doi.org/10.1038/s41598-025-15674-6
17. Deepak H a, Divya. G, Deepak Ramegowda, Keerthi Kumar M (2025), Butterfly Optimization based Feature Selection with Long-short Term Memory for Web Page Crawling, Conference: 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS), DOI:10.1109/ICICACS65178.2025.10969005