Received: 22 May 2025 | Revised: 12 August 2025 | Accepted: 22 October 2025 | Published online: 11 December 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1-18
DOI: 10.47852/bonviewAIA52026221
BON VIEW PUBLISHING

An Analytical Framework for Addressing Imbalance in Fake Review Detection Using Augmented Text and Swarm Optimized Classifier

Richa Gupta[1,2,*], **Indu Kashyap**[1], **and Vinita Jindal**[2]

[1] *School of Engineering & Technology, Manav Rachna International Institute of Research & Studies, India*

[2] *Department of Computer Science, University of Delhi, India*

Abstract: The rapid spread of online fake reviews threatens the consumers' trust, business reputation, and confidence in e-commerce platforms. Detecting fake reviews is challenging because distinguishing them from real reviews is difficult for humans. Researchers have employed various machine learning models to address fake review detection. However, publicly available datasets often suffer from severe class imbalance, with significantly fewer fake reviews than real ones. Previous research has struggled with this data imbalance problem, yielding biased results and thereby failing to detect fraud reviews efficiently. Traditionally, oversampling or undersampling techniques have been used to handle imbalance, which results in information loss and/or overfitting. To address this data imbalance problem, GlOBiL, a novel framework that uses GPT-2 with GloVe embedding and an optimized Bi-LSTM classifier, has been proposed. To optimize Bi-LSTM, a novel staggered particle swarm optimization (SPSO) algorithm has also been proposed. GlOBiL operates in three phases: Phase I generates contextually similar, synthetic fraudulent reviews via GPT-2 augmentation. Phase II processes this augmented dataset using GloVe embedding. Further, the Bi-LSTM classifier is optimized using the proposed SPSO. Finally, Phase III trains the optimized classifier and identifies the reviews as fake and real. Four benchmark review datasets—YelpZIP, YelpNYC, YelpCHI hotel, and YelpCHI restaurant—were used for experimentation. The results show that the proposed GlOBiL outperformed baseline methods and nine published approaches. Average accuracy increased to 95.37%, 96.49%, 97.67%, and 97.13%, respectively. Consequently, GlOBiL aids consumers and businesses in detecting fake reviews, enhancing trust, and supporting informed decision-making on e-commerce platforms.

Keywords: fake review detection, machine learning analytics, swarm optimization, text data augmentation, neural network training

1. Introduction

The prevalence of fake reviews is gaining popularity in the field of e-commerce and marketing. The primary factors contributing to this include the following: 1) the swift advancement of technology, enabling the generation of artificial content, and 2) the marketplace that has emerged around these artificial contents. This artificial content encompasses aspects such as the creation, detection, and mitigation of fakes. One of the most influential artificial marketing tactics is fraudulent product/service reviews. In simple language, they are known as "fake reviews," "fakes," "deceptive reviews," or "review fraud." Online product reviews serve as significant catalysts in shaping consumers' purchasing choices. According to a 2019 survey by BrightLocal, 76% of consumers placed the same level of trust in online reviews as they did in recommendations from friends [1].

Fake review detection (FRD) leverages natural language processing (NLP) to analyze sentiment, language patterns, contextual relevance, and metadata, uncovering anomalies that are indicative of deceptive or fraudulent reviews. Despite periodic warnings from both governments and websites, it remains challenging for the average consumer to differentiate between real and fake reviews. The inability to detect fake reviews places all stakeholders—e-commerce platforms, consumers, service providers, and businesses—at a considerable disadvantage. Protecting consumers from these misleading reviews is crucial. Because reviews greatly affect product rankings, businesses and service providers need to strengthen their defenses against unfair competition and take steps to safeguard their reputations [2]. Shen et al. [3] found that as the distribution of the ratings of a product or service deviates from what would typically be expected, it becomes more likely that these reviews have been manipulated or falsified. However, this is just one of the features.

Fake reviews can be either human-generated or machine-generated. Human-generated fake reviews are those that have been written by people who have not used the product or service but still write the review. In other cases, they might be given money or favors as an incentive to write fake reviews. Machine-generated reviews are the ones that use text-generation techniques to create fake reviews in bulk. These are usually bought by companies to boost their sales or downgrade their competitors [4]. Banerjee [5] investigated how much exaggeration is there in fake reviews and how a person perceives a review as either fake/fraudulent or real/authentic/genuine. They found that the fake reviews are not as exaggerated as they are thought to be and are similar to human-generated real reviews only. Thus, these fake reviews are undetected by the human eye. Past research has made significant strides in addressing the challenge of detecting

***Corresponding author:** Richa Gupta, School of Engineering & Technology, Manav Rachna International Institute of Research & Studies, India. Email: richa.gupta@keshav.du.ac.in

fraudulent reviews by employing various machine learning and deep learning algorithms [6]. These algorithms rely on a range of textual and behavioral features and demonstrate strong classification accuracy when trained on a large dataset that is evenly divided between fake and genuine reviews [7]. However, their effectiveness diminishes under two conditions: when the training data are limited or when one class heavily outweighs the other. The latter issue, known as the imbalanced dataset problem, serves as the motivation for this study.

Imbalanced datasets significantly affect FRD by skewing the model's learning process. When the number of fake reviews is much smaller than that of real reviews (or vice versa), the model tends to become biased toward the majority class, often predicting the majority class more frequently. This results in poor performance in detecting fake reviews because the model may overlook minority instances, leading to lower recall, precision, and overall accuracy for the minority class. There are labeled datasets available for FRD that are considered benchmarks. Four of the datasets used in this paper, namely, YelpNYC, YelpZIP, YelpCHI hotel, and YelpCHI restaurant by Rayana and Akoglu [8], exhibit a significant imbalance between fake and real reviews. This imbalance can lead to the development of inaccurate models and biased results. To address this issue, researchers have employed various strategies. Some choose to work exclusively with genuine reviews [9], and others utilize statistical techniques [10]. Clustering [11], oversampling [12], and cluster-based undersampling [13] in neural networks have been used to handle imbalances in data. However, these approaches come with their drawbacks, including the loss of valuable information, the introduction of noise, and uncertainty regarding the usefulness of the selected features. Some researchers introduce synthetic reviews to create a more balanced dataset [14]. This approach, although useful, may create unrealistic, semantically inaccurate, synthetic, fake reviews.

Hence, to overcome these drawbacks, a novel framework, GlOBiL, is proposed in this study, which generates contextual and semantically similar synthetic fake reviews using Generative Pre-trained Transformer-2 (GPT-2) and augments these reviews to the original dataset for classification. The classification of these reviews is conducted using optimized bidirectional long short-term memory (Bi-LSTM). This optimized Bi-LSTM leverages hyperparameter tuning through a novel staggered particle swarm optimization (SPSO) algorithm by finding the best configuration of the model parameters. The proposed framework results in improved accuracy and generalization capabilities on unseen data.

This study presents GlOBiL, a novel framework that addresses class imbalance in FRD by combining GPT-2 with GloVe embeddings and an optimized Bi-LSTM classifier, leveraging language generation to create balanced and contextually relevant data. It also introduces a new SPSO algorithm to optimize the classifier, employing a staggered update strategy for binary variables that enhances convergence speed and classification accuracy. Experimental results demonstrate that GlOBiL outperforms baseline models and nine existing methods across benchmark imbalanced datasets.

2. Related Work

Class imbalance is a significant challenge in the task of detecting fake reviews. This challenge arises because reviews obtained from online platforms lack clear indicators distinguishing them as genuine or fraudulent, even to human experts. Consequently, the distribution of fake and genuine reviews exhibits a severe skew, with a notably imbalanced ratio. Models trained on such imbalanced data tend to yield subpar and biased performance outcomes. Researchers have suggested various approaches to handle the class imbalance problem. This section presents a brief overview of the research conducted so far to handle the class imbalance problem.

2.1. Statistical sampling-based approaches

Several researchers have employed statistical approaches to address the challenge of data imbalance in FRD. Budhi et al. [15] proposed two dynamic random sampling techniques that combined undersampling and oversampling, which were applied to benchmark datasets such as YelpCHI, YelpNYC, and YelpZIP. Their findings revealed that oversampling provided minimal improvement in classification accuracy for smaller datasets and undersampling generally yielded better accuracy across most datasets, except when using the multilayer perceptron (MLP) method. However, undersampling uses only a limited portion of the majority class, and oversampling duplicates data from the minority class. This results in overfitting due to too much emphasis on the minority class. Singhal and Kashef [16] investigated ensemble sampling techniques to address data imbalance. Zhang et al. [17] utilized weighted latent Dirichlet allocation (LDA) and Kullback–Leibler (KL) divergence to uncover hidden topics within reviewer content and analyze similarities among reviewers favoring the fake reviews that are in the minority. To address dataset imbalance, they assigned distinct weights to fraudulent and authentic reviewers. Similarly, Zhang et al. [17] tackled the imbalance issue using expectation maximization and KL divergence. Yao et al. [18] addressed imbalanced data and reduced feature space using an ensemble approach that combined resampling and grid search. The grid search optimized the sampling ratio for each classifier. However, progressively increasing the sampling ratio for each resampling technique proved impractical for larger datasets. Cao et al. [19] employed multifeature learning and classification to train models on imbalanced Yelp datasets. Despite their effectiveness, these sampling methods have notable limitations, including the risk of overfitting caused by duplicating instances of the majority class (authentic reviews) and the potential loss of information caused by undersampling.

2.2. Crowd-sourced and SMOTE-based approaches

Some researchers have opted to introduce synthetic, fake reviews. This approach aims to create a more balanced dataset, mitigating the inherent data imbalance issue. Ott et al. [20] and Harris [21] tackled the issue of dataset imbalance by enlisting the help of crowdsourcing to introduce fake reviews, achieving a balanced dataset. Li et al. [22] employed random undersampling and borderline-SMOTE techniques, and Kumar et al. [23] used a modified version of the SMOTE algorithm to handle data imbalance. SMOTE generates synthetic minority samples by leveraging K-nearest neighbors instead of merely duplicating them. However, these artificially generated samples may not accurately reflect fake reviews and can lead to issues such as noise and class overlap.

Another approach by Gupta et al. [24] uses a Siamese neural network and Manhattan distance to determine the similarity between a pair of reviews. These review pairs may be real–real, real–fake, or fake–fake. It achieves good recall and precision over the fake review class, but it cannot achieve good results over the real review class.

2.3. Augmentation via deep learning and graph-based models

Some researchers use synthetic data augmentation for handling imbalance. Nayak et al. [25] detected opinion spam by generating synthetic fake reviews using sentiment classification, but their approach did not use the context or metadata of the text in reviews. Cheng et al. [26] proposed a GNN-based framework that captures information from different social network combinations in various subgraphs, addressing the issue of imbalanced data classification. However, the augmentation conducted by graph-based networks is susceptible to model collapse and data bias amplification. They may generate limited or repetitive variations of synthetic data, leading to a lack of diversity. Moreover, if the original dataset has a bias, the synthetic data may inherit and amplify those biases. Xu et al. [27] proposed a review-embedding framework so that reviews are encoded based on semantics and context. Then, they used attention loss, which placed greater emphasis on the difficult review samples in the imbalanced dataset. Luo et al. [28] developed a supervised probabilistic method utilizing linguistic, behavioral, and interrelationship features to address the imbalance challenge.

With the recent advancements in artificial intelligence and NLP, LLMs have been used as language generators that can produce human-like text based on inputs or prompts. Models such as BERT, RoBERTa, GPT, and Gemini generate coherent and contextually relevant text. Keya et al. [29] used BERT augmentation for fake news detection, and Atliha and Šešok [30] used BERT augmentation for image captioning. Similarly, GPT augmentation was used by Sawai et al. [31] for language translation, Cohen et al. [32] for enhancing social network hate detection, etc., and Mulla and Gharpure [33] used T5 to generate questions. However, data augmentation using language generators has not been used in the classification of fake and real reviews to the best of our knowledge.

2.4. Embedding of tokens

Machine learning algorithms require textual data to be converted into numerical representations. Common techniques include frequency-based methods (e.g., TF-IDF) and prediction-based methods (e.g., GloVe and BERT) [34]. Studies have shown that prediction-based methods are more effective in capturing word semantics [35]. Many researchers, including Chen and Yin [36], Ellaky et al. [37], and Muka and Mukala [38], have used GloVe embeddings for classification in fake news detection and FRD. Similarly, BERT embeddings have been used for the same by Khan and Shaikh [39] and Abduljaleel and Ali [40]. These studies show the importance of effective token embeddings in improving fake content detection models.

2.5. Swarm optimization for FRD

Several swarm intelligence-based techniques have been used for optimizing intrusion detection systems, fake news detection, and FRD. Ala'M et al. [41] used Harris hawks optimization to optimize the hyperparameters for spam review detection in multilingual review datasets. Deshai and Bhaskara Rao [42] employed adaptive particle swarm optimization (PSO) with transfer functions to convert continuous values into probabilities for effective future selection. Combined with recursive feature elimination, this approach improves the accuracy of fake profile detection. To address fake news detection, a study by Sirra et al. [43] used a deep quantum neural network optimized using cat swarm sea lion optimization to detect fake news on social media. These studies highlight the growing effectiveness of swarm optimization techniques in enhancing the accuracy and efficiency of fake content detection across various domains.

Hence, the proposed GlOBiL handles the class imbalance problem in FRD via language generator augmentation. It uses a similarity score for finding the best language generator, which can augment the dataset with contextually strong fake reviews. Furthermore, optimized Bi-LSTM, optimized using the proposed SPSO algorithm, has been used for training because it can look forward and backward, extracting potentially more relevant features from review text. Section 3 explains the methodology and architecture of the proposed GlOBiL framework.

3. Proposed Methodology

Data imbalance is a major challenge in detecting fake reviews. To address this issue, a novel framework, GlOBiL, is proposed, leveraging LLMs for their text generation abilities. GlOBiL combines GPT-2, GloVe embeddings, and an optimized Bi-LSTM classifier, operating in three phases (Figure 1). Phase I augments the original imbalanced dataset with synthetic fake reviews, Phase II embeds and optimizes the parameters of the Bi-LSTM via a novel SPSO algorithm, and Phase III classifies reviews as fake or real. The details of the three phases follow.

3.1. Phase I: augmentation of imbalanced datasets

As shown in Figure 1, Phase I augments the original dataset D with synthetically generated fake reviews F' to create a balanced, augmented dataset D', preventing classifier bias. Synthetic reviews from five LLMs—BERT, XLNet, RoBERTa, GPT-2, and Gemini Pro 1.0—were evaluated based on a semantic similarity score. GPT-2 achieved the highest similarity to original fake reviews. GPT-2 was utilized in its pretrained form, adopting a contextual augmentation strategy rather than fine-tuning or prompt-based generation. Each original fake review was slightly modified using GPT-2's contextual embeddings, generating domain-relevant synthetic text that remains close in semantics and style to the original dataset. This approach:

1) eliminated the cost of fine-tuning LLMs,
2) preserved domain-specific style, and
3) avoided off-topic or syntactic inconsistency. The resulting F' was combined with human-written real R and fake F reviews to form D' (Figure 2).

Existing augmentation methods used for fake news detection [29], image captioning [30], language translation [31], etc., either 1) generate full reviews from scratch using prompt-based generation [14] or 2) require computationally expensive fine-tuning [44], both causing domain drift and high computational costs. In contrast, the proposed contextual augmentation strategy is lightweight, controlled, and domain-relevant and avoids fine-tuning, making it well suited for FRD [45]. This approach, to the best of our knowledge, is novel for FRD. Moreover, an empirical evaluation of five LLMs, namely, BERT, XLNet, RoBERTa, GPT-2, and Gemini Pro 1.0, was conducted, and GPT-2 was chosen based on semantic similarity scores, rather than relying on assumed LLM quality.

Devlin et al. [46] introduced BERT in October 2018 for various NLP tasks. It processes text bidirectionally, understanding the meaning and relationships between words. Hence, it was used as a solid baseline for the Phase I experiment. XLNet, introduced by Yang et al. [47] in 2019 and improved BERT by generating text autoregressively, has been widely used for text completion and creative writing. Hence, it was used as an improvised baseline for Phase I. RoBERTa, introduced by Liu et al. [48], is primarily an optimized variant of BERT that focuses on pretraining without word masking, yielding better performance and enabling fine-tuned text generation.

GPT-2, introduced by Radford et al. [49] in late 2019, is a causal transformer pretrained on large-scale language modeling to predict the next word based on context. Its ability to generate coherent, human-like text made it suitable for producing synthetic fake reviews. Released in December 2023, Gemini encompasses Gemini Ultra, Gemini Pro, and Gemini Nano, tailored for "highly complex," "a wide range," and "on-device" tasks, respectively [50]. Gemini Pro 1.0 was selected for its accessibility and comprehensiveness.

These five LLMs share a transformer-based architecture, enabling them to capture long-range dependencies in text and generate coherent, contextually appropriate language. Trained on vast text corpora, they learn intricate language patterns, grammar, and knowledge from diverse domains. To the best of our knowledge, data augmentation using language generators has not previously been applied to FRD. Although these models share the transformer architecture and are pretrained on large text corpora, they differ in their training objectives, data,

Figure 1
Architecture of the proposed framework GlOBiL

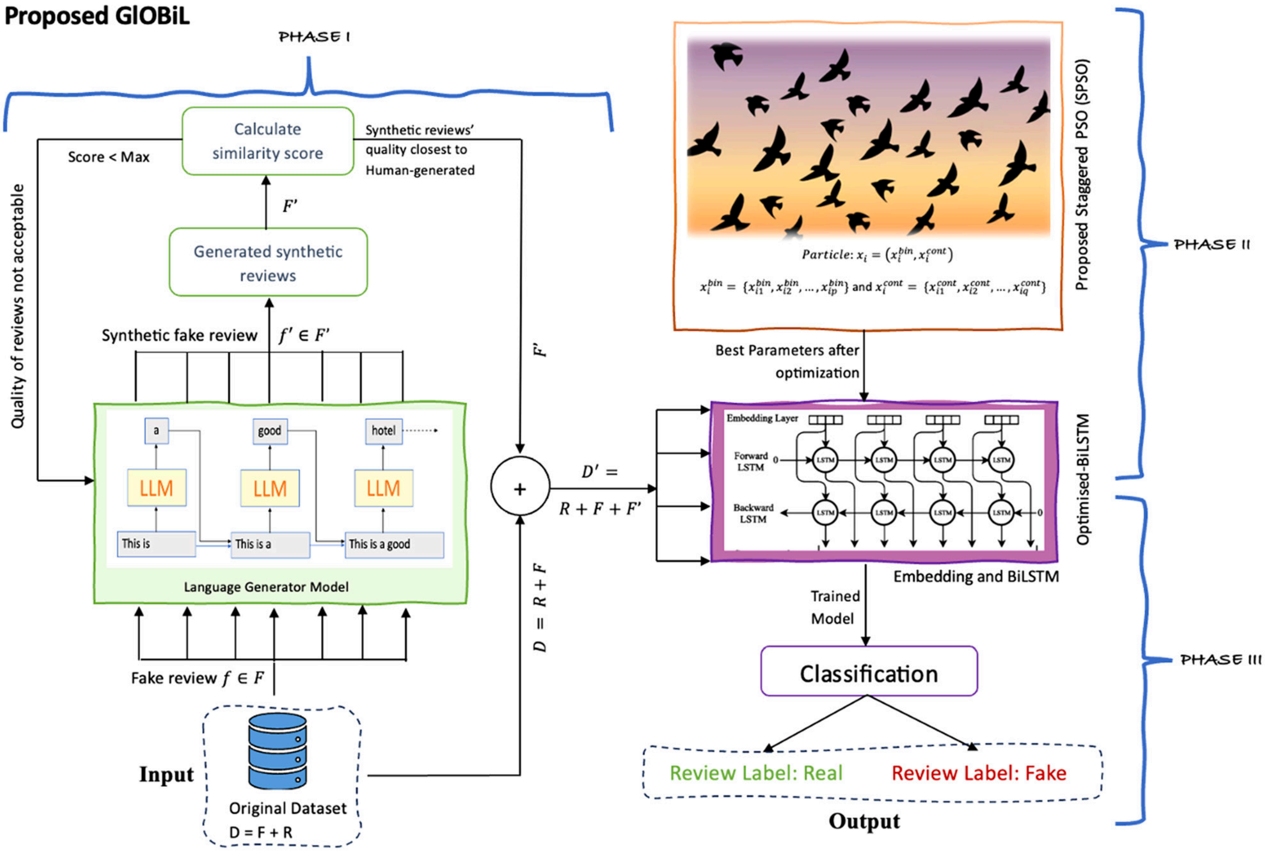


Figure 2
Pseudocode for Phase I of the proposed framework GlOBiL

ALGORITHM 1: Phase I Augmentation of the imbalanced dataset

Input : $Imbalanced\ Dataset\ D = \{fake\ reviews\ F\ (minority) +\ Real\ reviews\ R\ (majority\ class)\}$
Output: $Balanced\ dataset\ D' = \{fake\ reviews\ F + generated\ fake\ reviews\ F' + Real\ reviews\ R\}$

```
Function augmentation (D):
    for each LLM_i, i = 1 to 5
        real R ← review with label = 1
        fake F ← review with label = −1
        F'_i ← gen_rev(LLM_i, F)
    end
    # To find similarity score Score_sim
    Score_sim^max ← 0, bestLLM ← 0
    for i = 1 to 5
        Remove duplicates from subset S_i = {F + F'_i}, F ∈ D
        S'_i = S_i − duplicates
        simlarity computation of each pair [f, f'_i], where f ∈ F and f'_i ∈ F'_i
        compute average of above as Score_sim^i
        if Score_sim^i > Score_sim^max
            Score_sim^max ← Score_sim^i
            bestLLM ← i
    end
    return new dataset D' = {F + F'_bestLLM + R}, where n_F + n_F' = n_R
end
```

size, bidirectionality, and pretraining methods, which influence their performance and suitability for FRD.

In Phase I, synthetic fake reviews were generated using each of the five LLMs, finding whether the synthetic reviews are of contextually good quality or not and then appending the best generated synthetic fake reviews to the original dataset. In addition, data preprocessing, typically performed to ensure the consistency and quality of the dataset, such as removal of punctuation, diacritics, and whitespaces, was applied to create the final augmented D', which was then passed on to Phase II for further processing.

3.2. Phase II: embedding and optimization of the Bi-LSTM classifier using novel SPSO

Phase II involves the following:

1) embedding the augmented dataset D' as the first step, ensuring that the data are transformed into a suitable format for training
2) optimization of the Bi-LSTM model via the proposed SPSO as the second step, focusing on optimizing the Bi-LSTM parameters using novel SPSO.

3.2.1. Embedding

The first step involves passing the augmented dataset D' through an embedding layer. Global Vectors for Word Representation (GloVe) was chosen to represent each review as a sequence of numerical vectors. Because word embeddings strongly influence FRD performance, two popular techniques—GloVe and BERT—were compared. GloVe uses global word–word co-occurrence statistics, whereas BERT captures contextual information. Experiments showed that GloVe outperformed BERT for FRD. Being widely used in NLP classification tasks, a GloVe embedding layer was added before passing the augmented dataset to the Bi-LSTM classifier.

3.2.2. Optimizing Bi-LSTM using novel SPSO

In this second step, Bi-LSTM, introduced by Schuster and Paliwal [51] in 1997, is optimized using a novel SPSO algorithm. Bi-LSTM processes input sequences bidirectionally, improving accuracy in tasks such as FRD, and is used here as the classifier. As a recurrent neural network (RNN) and nongenerative language model, it differs from GPT-2 in both architecture and tokenizer. Unlike GPT-2, Bi-LSTM is inherently bidirectional. This distinction becomes significant because synthetic review creation in Phase I used GPT-2. To ensure independence from GPT-2's generative style and enhance classification accuracy, GlOBiL leverages Bi-LSTM's strengths while minimizing GPT-2's influence.

PSO: PSO, which was introduced by Kennedy and Eberhart in 1995, is a metaheuristic inspired by birds' foraging behavior, where particles exchange positional information to locate an optimal solution. Each particle's position x_i represents a potential solution, and the best solution found corresponds to the optimal source. Each particle maintains a memory to store its best-found position and the global best position among all particles in the swarm. On the basis of this stored information and the particle's velocity v_i, it updates its position x_i (t+1) in the search space. PSO balances exploration and exploitation and is simple and efficient for continuous domains but struggles with optimizing discrete or categorical parameters.

Binary PSO: it is a variant of the PSO algorithm designed to optimize problems with binary decision variables. In this, particle positions represent binary solutions, with updates determined probabilistically through a sigmoid-transformed velocity, influencing whether a bit flips or remains unchanged. However, it performs less efficiently in continuous or real-valued domains, leading to slower convergence in them.

Proposed SPSO: SPSO offers distinct advantages over using binary or standard PSO alone. It allows parallel optimization of binary parameters (e.g., whether to apply dropout (yes/no)) and continuous hyperparameters (e.g., learning rate, dropout rate, and LSTM unit size) without converting one type into another, avoiding unnecessary complexity. Dedicated mechanisms for handling each type of hyperparameter lead to more efficient and accurate search due to faster convergence across the hyperparameter space. This results in improved optimization efficiency. However, a key limitation of this approach is its tendency to be trapped in local optima. In other words, the continuous variables may not undergo sufficient exploration or refinement within the search space because their convergence typically requires more iterations than binary variables [52]. To overcome this problem, binary variables are initialized to fixed values (0 or 1) and held static during early iterations to minimize frequent state transitions. Updates are deferred until the kth iteration, determined experimentally. This staggered update process, shown in Figure 3, enables continuous and binary parameters to evolve independently within their respective search spaces, allowing for more effective convergence. Hence, SPSO adapts to complex search spaces with both discontinuities (binary parameters) and smooth gradients (continuous parameters).

The SPSO algorithm starts by optimizing the continuous parameters in every iteration. It waits until after the kth iteration to begin optimizing the binary parameter. The penalty is used to include binary optimization gradually, making it less important early on and more important later. Once both optimizations are complete, the best parameters are used to train the final model and test it. GlOBiL uses the proposed novel SPSO to optimize Bi-LSTM. The binary and continuous parameters are updated separately with the following update rules and the fitness function, which is used to guide the optimization. SPSO can be represented mathematically using the following equations:

Each particle i consists of two parts:

1) a binary vector $x_i^{bin} \in \{0,1\}^p$, where p is the no. of binary parameters
2) a continuous vector $x_i^{cont} \in \Re^q$, where q is the number of continuous hyperparameters.

Thus, a PSO particle x_i can be written as given in Equation (1).

$$x_i = \left(x_i^{bin}, x_i^{cont}\right), \tag{1}$$

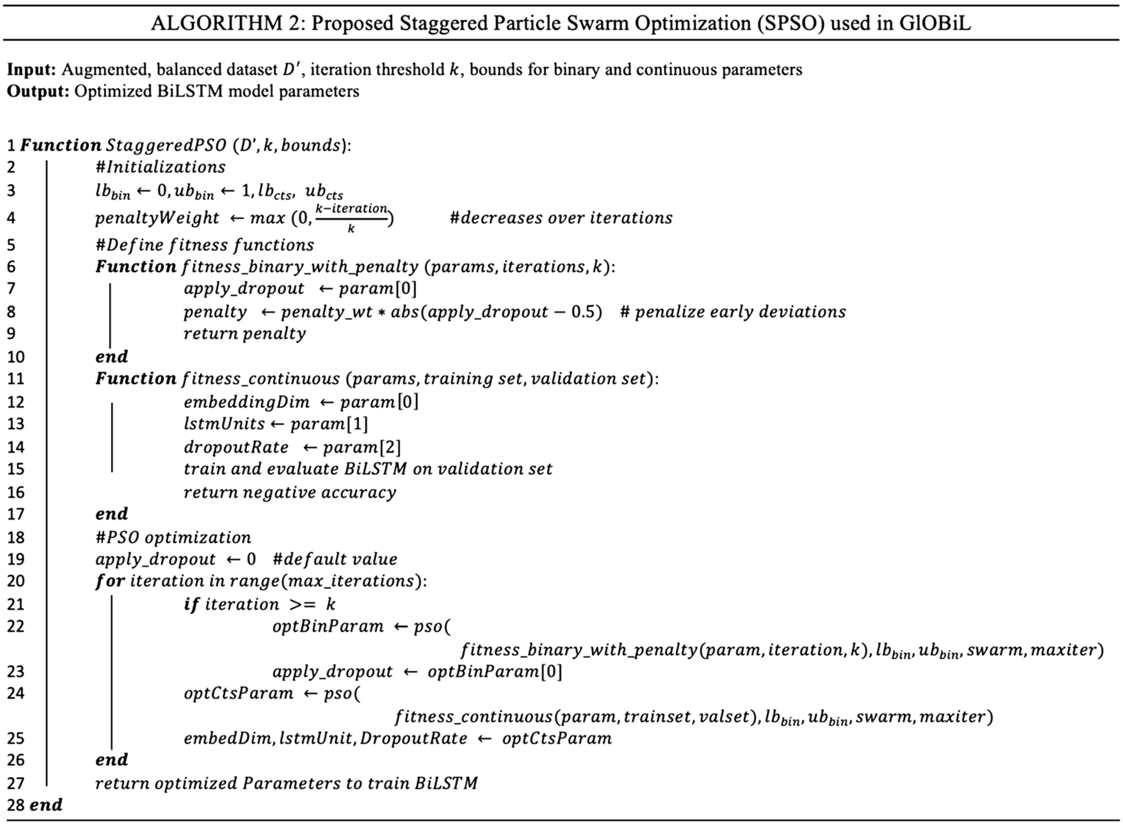
where $x_i^{bin} = \left\{x_{i1}^{bin}, x_{i2}^{bin}, \ldots, x_{ip}^{bin}\right\}$ is the binary component and $x_i^{cont} = \left\{x_{i1}^{cont}, x_{i2}^{cont}, \ldots, x_{iq}^{cont}\right\}$ is the continuous component. Each particle x_{ij} also has binary velocity v_{ij}^{bin} and continuous velocity v_{ij}^{cont}. The update equations for the binary and continuous velocities are given in Equations (2) and (3), respectively.

$$\begin{aligned} v_{ij}^{bin}(t+1) = w \cdot v_{ij}^{bin}(t) + c_1 \cdot r_1 \cdot \left(p_{ij}^{binbest} - x_{ij}^{bin}(t)\right) \\ + c_2 \cdot r_2 \cdot \left(g_j^{binbest} - x_{ij}^{bin}(t)\right), \end{aligned} \tag{2}$$

$$\begin{aligned} v_{ij}^{cont}(t+1) = w \cdot v_{ij}^{cont}(t) + c_1 \cdot r_1 \cdot \left(p_{ij}^{contbest} - x_{ij}^{cont}(t)\right) \\ + c_2 \cdot r_2 \cdot \left(g_j^{contbest} - x_{ij}^{cont}(t)\right), \end{aligned} \tag{3}$$

where w is the inertia weight controlling the effect of previous velocities; c_1, c_2 are acceleration coefficients; r_1, r_2 are random

Figure 3
Pseudocode for the proposed SPSO algorithm



ALGORITHM 2: Proposed Staggered Particle Swarm Optimization (SPSO) used in GlOBiL

Input: Augmented, balanced dataset D', iteration threshold k, bounds for binary and continuous parameters
Output: Optimized BiLSTM model parameters

```
Function StaggeredPSO ($D', k, bounds$):
    #Initializations
    $lb_{bin} \leftarrow 0, ub_{bin} \leftarrow 1, lb_{cts}, ub_{cts}$
    $penaltyWeight \leftarrow max\ (0, \frac{k-iteration}{k})$        #decreases over iterations
    #Define fitness functions
    Function fitness_binary_with_penalty (params, iterations, k):
        apply_dropout  ← param[0]
        penalty  ← penalty_wt * abs(apply_dropout − 0.5)   # penalize early deviations
        return penalty
    end
    Function fitness_continuous (params, training set, validation set):
        embeddingDim  ← param[0]
        lstmUnits ← param[1]
        dropoutRate  ← param[2]
        train and evaluate BiLSTM on validation set
        return negative accuracy
    end
    #PSO optimization
    apply_dropout  ← 0   #default value
    for iteration in range(max_iterations):
        if iteration  >=  k
            optBinParam  ← pso(
                        fitness_binary_with_penalty(param, iteration, k), $lb_{bin}, ub_{bin}$, swarm, maxiter)
            apply_dropout  ←  optBinParam[0]
        optCtsParam  ← pso(
                    fitness_continuous(param, trainset, valset), $lb_{bin}, ub_{bin}$, swarm, maxiter)
        embedDim, lstmUnit, DropoutRate  ←  optCtsParam
    end
    return optimized Parameters to train BiLSTM
end
```

numbers uniformly distributed in [0,1]; $p_{ij}^{binbest}$ and $p_{ij}^{contbest}$ are the personal best binary and continuous positions of particle i, respectively; and $g_j^{binbest}$ and $g_j^{contbest}$ are the global best binary and continuous positions of particle i, respectively. The update equation for binary parameter positions is given in Equation (4).

$$x_{ij}^{bin}(t+1) = \begin{cases} 1, & \text{if } r < \mathbb{S}\left(v_{ij}^{bin}(t+1)\right) \\ 0, & \text{otherwise} \end{cases}, \qquad (4)$$

where $x_{ij}^{bin}(t+1)$ is the new position, r is a random number between 0 and 1, and $\mathbb{S}$ is the sigmoid transformation given by Equation (5).

$$\mathbb{S}\left(v_{ij}^{bin}\right) = \frac{1}{1+e^{-v_{ij}^{bin}(t+1)}}. \qquad (5)$$

The update equation for continuous parameter positions is given in Equation (6).

$$x_{ij}^{cont}(t+1) = x_{ij}^{cont}(t) + v_{ij}^{cont}(t+1). \qquad (6)$$

This update ensures that each continuous parameter changes in response to a change in velocity. After the above position update, the continuous values are restricted to remain within predefined bounds $[x_{j,min}, x_{j,max}]$, as represented in Equation (7).

$$x_{ij}^{cont}(t+1) = \min\left(\max\left(x_{ij}^{cont}(t+1), x_{j,min}\right), x_{j,max}\right). \qquad (7)$$

After the update, fitness function $f(x_i)$ for each particle i is evaluated, which is the validation accuracy of the model trained with the corresponding hyperparameters. After each fitness evaluation, the personal best p_i^{best} and global best g^{best} of particle i are updated, as shown in Equations (8) and (9).

$$p_i^{best} = x_i \quad \text{if} \quad f(x_i) < f\left(p_i^{best}\right), \qquad (8)$$

$$g^{best} = x_i \quad \text{if} \quad f(x_i) < f\left(g^{best}\right). \qquad (9)$$

The next subsection explains the last phase of the proposed GlOBiL framework.

3.3. Phase III: classification and testing

Phase III classifies the reviews as fraudulent or genuine. It trains and tests the proposed framework. The augmented, preprocessed, and embedded dataset D' is fed into the optimized Bi-LSTM for training. It is a deep learning classifier selected for its capability to capture the semantics and dependencies in both forward and backward directions. The rationale behind the selection of Bi-LSTM as the classifier was discussed in the previous subsection. The "Adam" optimizer was used during the framework's compilation due to its simplicity, memory efficiency, and effectiveness with large datasets; it was used to minimize the "binary cross-entropy" function, which generates high values for inaccurate predictions and low values for accurate ones. Specifically designed for binary classification tasks, binary cross-entropy measures the difference between true labels and predicted probabilities. It is represented by Equation (10).

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1-y_i) \cdot \log(1-p(y_i)). \qquad (10)$$

Here, $H_p(q)$ represents the cross-entropy loss H between two probability distributions p (true labels) and q (predicted probabilities). "yi" represents the true label, with 1 indicating a genuine review and 0 indicating a fake one, and "p(yi)" signifies the predicted probability of the review as determined by the classification model. The next section details the experiments conducted using publicly available benchmark Yelp datasets.

4. Experimental Setup

The proposed framework, GlOBiL, has been designed to handle the challenge of data imbalance in FRD. The framework has been designed and developed using an Apple M2 processor on macOS Ventura version 13.3. Python version 3.10.12 with Tensorflow and Keras APIs were used for experimentation. This section provides a comprehensive explanation of the datasets used and the technical aspects of all three phases involved in the proposed framework's experiments.

4.1. Data collection and preparation

GlOBiL has been tested on four benchmark datasets collected by Rayana and Akoglu [8], namely, YelpZIP, YelpNYC, YelpCHI hotel, and YelpCHI restaurant. These datasets depict a significant level of imbalance between the real and fake reviews, the statistics of which are given in Table 1. Yelp is known for its extensive and trustworthy reviews for businesses, particularly restaurants, cafes, hotels, and other local services [53, 54]. Labeling of reviews as fake or authentic has been performed using Yelp FRD algorithms and is considered free from human error. Although Yelp's filtering mechanism is not flawless, research has shown that it delivers accurate results. The filtering process is considered both effective and reliable in identifying questionable reviews [55]. Table 1 demonstrates a remarkably low percentage of fake reviews, indicating a significant imbalance in the datasets. In these situations, traditional machine learning algorithms often generate biased results that favor the majority class. As a result, these datasets were chosen for this experiment to explore their pronounced imbalance and to explore potential solutions for the imbalanced data issue.

4.2. Experiments in Phase I: augmentation of imbalanced dataset and data preprocessing

The first phase leverages five LLMs to generate synthetic reviews and selects the one that generates reviews semantically and contextually closest to the human-generated fake reviews in the Yelp datasets. The pseudocode for this is given in Figure 2. Each of the five LLMs generated $f' \in F'$ for each $f \in F$. A sample of the same is given in Table 2. As illustrated, scanning the generated reviews manually reveals that all five models produce similar outputs, making it difficult to identify the best generator. Therefore, a metric was developed to determine which model generated fraudulent reviews most closely resembling human-written ones.

To achieve this, a "similarity score," i.e., word similarity measured on a scale from 0 to 1 reflecting the degree of semantic closeness between the two reviews, was calculated by comparing word vectors within a vector space. SpaCy, a high-performance NLP library, was used due to its efficient tokenization, word embeddings, and linguistic features, making it well suited for computing review similarity. For the analysis, all possible pairs of human-generated and machine-generated fake reviews were formed as [f, f′], where f represents a review from the human-generated fake subset F and f′ represents a review from the machine-generated fake subset F′.

A similarity score was computed for each pair using SpaCy. Figure 4 shows a sample result, where a score of 0.944117 indicates 94.41% semantic and contextual similarity between the synthetic and human-generated review pair. After computing the similarity for all pairs, the average similarity score for each language generator was determined, as presented in Table 3. It is observed that GPT-2 achieved the highest mean similarity score among the five language generators. GPT-2-generated fakes were of high quality and contextually similar as the original dataset. Hence, it is considered the best among the five and is used for creating the augmented dataset D' for identifying fake reviews.

Hence, GPT-2-generated fake reviews F' were added to the original dataset D = F + R, resulting in an augmented and balanced dataset D′ = F + F′ + R. This process resulted in an equal number of fake and real reviews, achieving a balanced dataset for further analysis. Specifically, F + F′ = R, ensuring parity between fake and real reviews. In addition, augmented datasets were created using other models, including BERT, RoBERTa, XLNet, Gemini Pro, and the Synthetic Minority Oversampling Technique (SMOTE), to compare results. However, because BERT has a maximum token length of 512 and RoBERTa and XLNet are based on BERT, only reviews with 512 tokens or fewer were selected. These selected reviews were then fed into the respective language generator, and the generated fake reviews were appended to the original dataset.

This augmented dataset is then cleaned and preprocessed for the next phases. Data preprocessing is a crucial step in preparing the raw input data for building a machine learning model. Preprocessing of these four datasets is conducted by removing any missing or null values in the dataset, eliminating punctuations because they are irrelevant to text classification, and finally, converting all review text to lowercase to simplify parsing. These refined, augmented corpora are now used for the next phases of GlOBiL.

4.3. Experiments in Phase II: embedding of augmented dataset and optimization of Bi-LSTM using SPSO

Phase II involves experimentation in two steps—embedding of the augmented datasets and optimization of Bi-LSTM using the proposed SPSO algorithm.

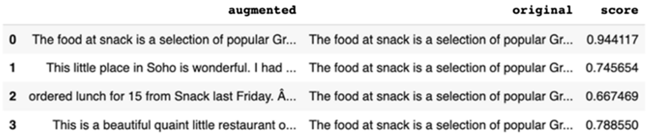
Table 1
Descriptions of benchmark datasets used for experimentation

	Fake reviews		Real reviews		
Dataset name	***Number***	***Percentage***	***Number***	***Percentage***	**Total reviews**
YelpZIP	80,466	13.22	528.132	86.78	608,598
YelpNYC	36,885	10.27	322,167	89.73	359,052
YelpCHI restaurant	8141	13.23	53,400	86.77	61,541
YelpCHI hotel	778	13.29	5076	86.71	5854

Table 2
Sample of synthetic reviews generated by the five language generators

Human-generated fake review	LLM	Synthetic (similar) fake review
The food at Snack is a selection of popular Greek dishes. The appetizer tray is good as is the Greek salad. We were underwhelmed with the main courses. There are 4–5 tables here so it's sometimes hard to get seated [8].	BERT	At Snack, you'll encounter a delightful array of popular Greek dishes to tantalize your taste buds. The appetizer platter stands out, and the Greek salad is a true treat. However, we found ourselves somewhat underwhelmed with the main courses. Do keep in mind that this charming establishment boasts only a handful of tables, making it occasionally tricky to secure a seat. At Snack, you'll find an enticing variety of well-loved Greek cuisine. The appetizer platter is truly delightful, and the Greek salad is a real standout. However, our enthusiasm waned when it came to the main courses. It's worth noting that this cozy establishment only accommodates a handful of tables, which can occasionally pose a challenge when trying to secure a seat.
	XLNet	Snack offers a range of popular Greek dishes that are sure to satisfy your cravings. The appetizer platter is quite enjoyable, and the Greek salad is a real treat. However, when it comes to the main courses, our expectations weren't entirely met. It's worth noting that this cozy spot has only a handful of tables, which can occasionally make finding a seat a bit challenging.
	GPT-2	At Snack, you'll find a delightful variety of well-known Greek dishes to tantalize your taste buds. The appetizer platter is a standout, and the Greek salad is absolutely delicious. However, we must admit that the main courses left us a bit underwhelmed. It's worth noting that this charming establishment only has a handful of tables, which can make it a bit tricky to secure a seat at times.
	Gemini Pro 1.0	The food at Snack is a delicious assortment of traditional Greek cuisine. The appetizer platter was a feast for the senses, and the Greek salad was fresh and flavorful. However, we were slightly disappointed with the main courses, which didn't quite live up to our expectations. With only a few tables available, it can be challenging to find a seat during peak times.
This little place in Soho is wonderful. I had a lamb sandwich and a glass of wine. The price shocked me for how small the serving was, but then again, this is Soho. The staff can be a little snotty and rude, but the food is great, just don't expect world-class service [8]	BERT	This cozy spot in the heart of Soho is truly delightful. I indulged in a delectable lamb sandwich paired with a glass of wine. While I was a bit surprised by the price given the modest portion size, it's worth remembering that we are in the heart of Soho. The staff here can be a tad aloof and occasionally impolite, but the food is superb. Just keep in mind that you're here for the culinary experience, not necessarily world-class service.
	RoBERTa	This charming little gem nestled in Soho is an absolute delight. I savored a mouth-watering lamb sandwich along with a glass of fine wine. Although I was taken aback by the cost considering the somewhat modest serving size, it's important to keep in mind that Soho is known for its upscale offerings. The service staff may come across as a bit reserved or brusque at times, but the culinary experience here is truly exceptional. Just bear in mind that your visit is primarily about savoring remarkable cuisine, rather than anticipating top-tier service
	XLNet	This quaint little spot in the heart of Soho is absolutely wonderful. I indulged in a delicious lamb sandwich accompanied by a fine glass of wine. While I was a bit surprised by the pricing given the relatively modest portion size, it's important to consider that Soho is renowned for its upscale offerings. The service staff, it must be noted, can be somewhat aloof and occasionally brusque, but the culinary experience here is truly exceptional. Just remember that your visit is primarily about savoring outstanding cuisine, rather than expecting top-notch service
	GPT-2	This charming little place in the heart of Soho is absolutely wonderful. I decided to indulge in a mouth-watering lamb sandwich paired with a delightful glass of wine. While the price did give me a bit of a shock, considering the somewhat modest portion size, it's worth noting that Soho is known for its upscale dining scene. The service staff can come across as a bit reserved and occasionally brusque, but the food more than makes up for it. Just keep in mind that your visit here is all about savoring exceptional cuisine, rather than expecting world-class service.
	Gemini Pro 1.0	This charming little spot in Soho is a hidden gem. I indulged in a succulent lamb sandwich and a glass of wine. While the portion size might seem a bit on the small side for the price, it's worth noting that this is Soho after all. The food is absolutely delicious, but don't expect top-notch service from the staff, who can sometimes come across as a bit aloof.

Figure 4
Similarity score of a sample pair of reviews



	augmented	original	score
0	The food at snack is a selection of popular Gr...	The food at snack is a selection of popular Gr...	0.944117
1	This little place in Soho is wonderful. I had ...	The food at snack is a selection of popular Gr...	0.745654
2	ordered lunch for 15 from Snack last Friday. Â...	The food at snack is a selection of popular Gr...	0.667469
3	This is a beautiful quaint little restaurant o...	The food at snack is a selection of popular Gr...	0.788550

Table 3
Mean similarity score of each LLM

Language generator	Mean similarity score
BERT	0.5869451070090201
XLNet	0.5657289053531175
RoBERTa	0.5675505410715225
GPT-2	**0.5932356237075136**
Gemini Pro 1.0	0.5891045902311276

Step 1: embedding of the augmented dataset. The choice of word embeddings can significantly affect the performance of any model. GloVe and BERT are two popular word embedding techniques used by researchers. GloVe uses global word–word co-occurrence statistics to learn word embeddings, whereas BERT captures contextual information. In this study, both embedding layers are experimented upon, and the results presented in Table 4 show that GloVe embedding gave better results than BERT embedding. Although BERT gives contextual embeddings, in the case of FRD, GloVe gave better embeddings and is widely used for classification problems in NLP. The experiment was conducted for all four datasets using five-fold cross-validation. It is important to note that Bi-LSTM was not optimized for selecting the embedding layer.

It was observed that GloVe embedding gave better results in terms of accuracy and F1 score. This observation may be explained by the nature of fake reviews, which often contain repetitive phrasing, exaggerated sentiment, or templated expressions [56]. GloVe, being a frequency-based embedding method, captures global co-occurrence patterns that are effective in detecting such regularities. In contrast, BERT's contextual representations may dilute these global patterns or introduce unnecessary noise, especially when the input reviews are relatively short or lack sufficient context. Furthermore, BERT embeddings are highly dimensional and computationally heavier, which could require more extensive fine-tuning to yield optimal results for this task. Because the focus here was on the effectiveness of data augmentation in FRD, GloVe was chosen as a more efficient choice. This aligns with prior studies in FRD [36, 38], which also report superior performance with GloVe embeddings in classification tasks.

To visually support this explanation, four t-SNE graphs illustrating the difference in clustering between GloVe and BERT embeddings on a subset of fake vs. real reviews are plotted. As shown in Figure 5, GloVe better separates fake and real reviews into distinct clusters, whereas BERT embeddings display some class overlap, likely due to contextual noise. Hence, the GloVe embedding layer was added before passing the augmented dataset to the Bi-LSTM layer.

Step 2: optimization of Bi-LSTM. This step involves computing the average accuracy and other performance metrics through classification 1) without optimizing Bi-LSTM and 2) after optimizing Bi-LSTM using the proposed SPSO. The results of these experiments, presented in Table 5, highlight the performance improvement achieved by applying SPSO optimization to Bi-LSTM across various datasets. Optimized Bi-LSTM consistently outperforms its nonoptimized counterpart due to faster and more reliable convergence achieved by balancing exploration and exploitation of binary and continuous parameters.

It is important to clarify that although hybrid binary-continuous PSO approaches do exist [57], the proposed SPSO introduces a distinct and theoretically grounded contribution through its staggered update mechanism. Unlike conventional HPSO techniques that update both binary and continuous parameters simultaneously, SPSO first optimizes the continuous variables, activating binary parameters only after k iterations. This delay mitigates the risk of premature convergence often observed in HPSO due to early binary updates. SPSO's use of time-dependent penalty ensures a smooth transition into the binary search phase. This not only improves the balance between exploration and exploitation but also helps in avoiding premature convergence and poor local optima. Theoretically, SPSO's staggered design allows for finer control over search complexity, making it especially suitable for the proposed framework. Empirically, this mechanism shows improved convergence behavior as shown in Figure 6. Because the fitness function is the validation accuracy, higher values correspond to better-performing solutions.

Figure 6 presents the convergence behavior of three algorithms, namely, PSO, hybrid PSO (binary-continuous), and the proposed SPSO, over 20 iterations. To ensure fairness and consistency in comparison, the initial population across all algorithms was identically initialized using the same initial configuration and seed. It can be seen across datasets that PSO barely improves and quickly gets trapped in a local optimum. Hybrid PSO converges quickly because of binary optimization but plateaus prematurely. In contrast, SPSO converges more steadily and ultimately reaches the best solution. These findings support SPSO as the preferred choice for GlOBiL.

Table 4
Comparison of GloVe and BERT embeddings on the performance of the augmented dataset

Dataset	Embedding	Accuracy (%)	Precision	Recall	F1 score	AUC score
YelpZIP	GloVe	**94.35**	**0.98**	**0.90**	**0.94**	**0.9436**
	BERT	93.25	0.94	0.93	0.94	0.9320
YelpNYC	GloVe	**94.33**	**0.97**	**0.91**	**0.94**	**0.9433**
	BERT	93.66	0.95	0.91	0.93	0.9348
YelpCHI	GloVe	**95.79**	**0.99**	**0.91**	**0.95**	**0.9580**
	hotel	92.97	0.93	0.90	0.92	0.9317
YelpCHI restaurant	GloVe	**95.93**	**0.99**	**0.92**	**0.95**	**0.9597**
	BERT	94.52	0.95	0.93	0.94	0.9458

Figure 5
t-SNE visualizations of GloVe and BERT embeddings across the four Yelp datasets

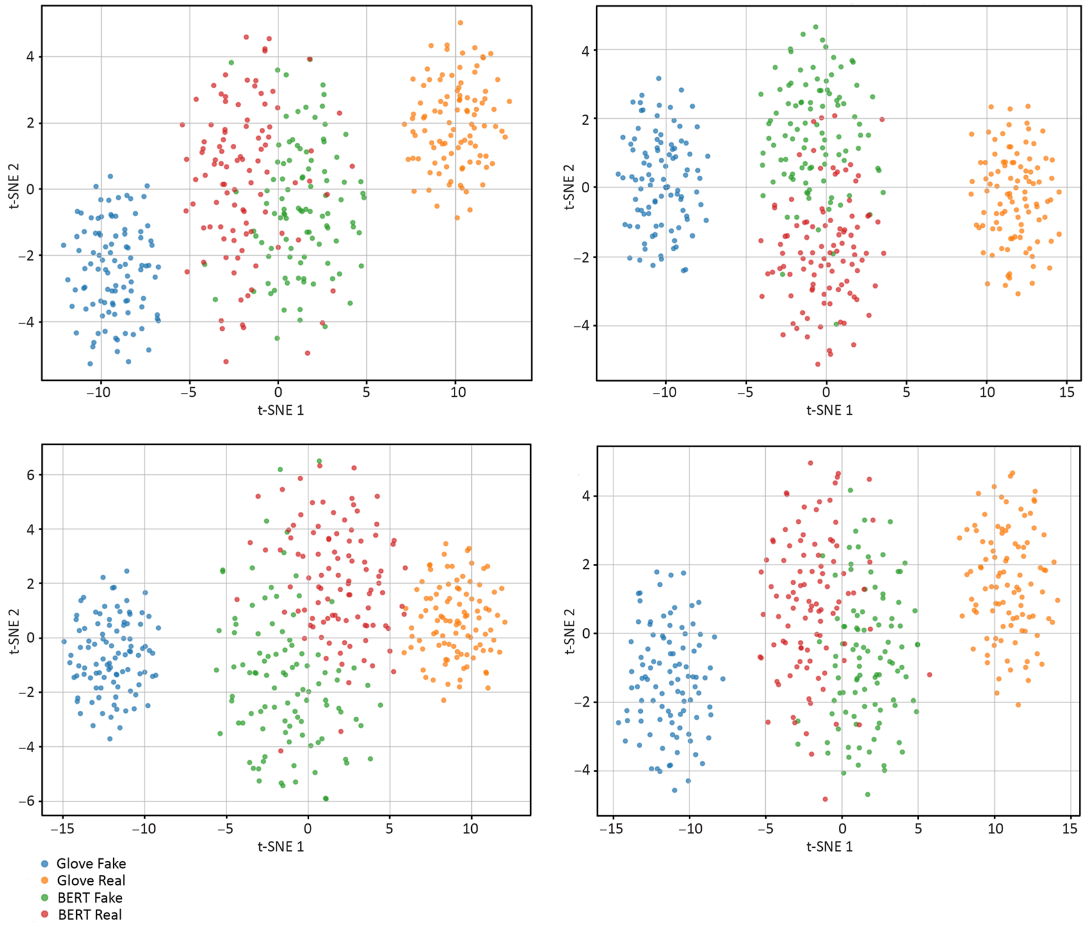


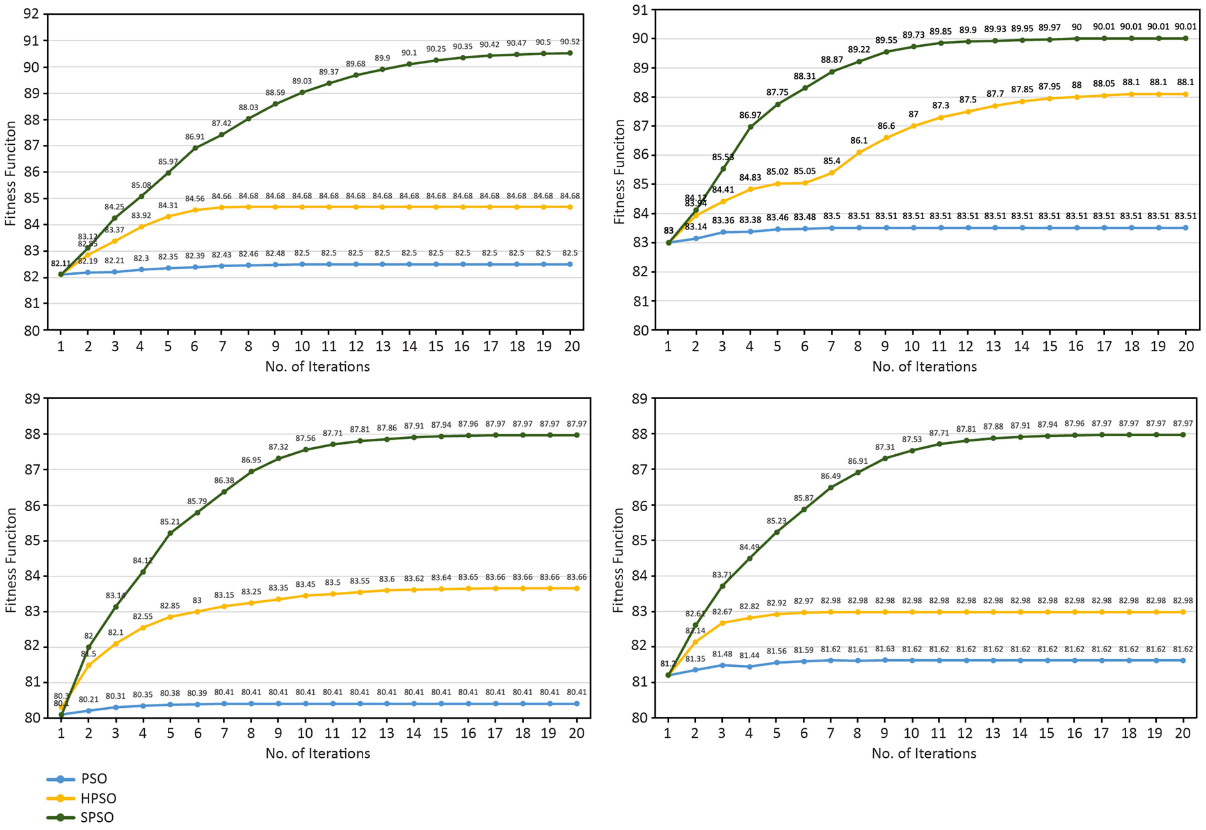
Table 5
Performance comparison of various optimized and nonoptimized Bi-LSTM classifications

Dataset	Classification using	Accuracy (%)	Precision	Recall	F1 score	AUC
YelpZIP	(a) Bi-LSTM w/out optimization	94.35	0.98	0.90	0.94	0.9436
	(d) Bi-LSTM w/ SPSO (proposed GlOBiL)	**95.37**	**0.99**	**0.91**	**0.95**	**0.9538**
YelpNYC	(a) Bi-LSTM w/out optimization	94.33	0.97	0.91	0.94	0.9433
	(d) Bi-LSTM w/ SPSO (proposed GlOBiL)	**96.49**	**0.99**	**0.93**	**0.96**	**0.9648**
YelpCHI hotel	(a) Bi-LSTM w/out optimization	95.79	0.99	0.91	0.95	0.9580
	(d) Bi-LSTM w/ SPSO (proposed GlOBiL)	**97.67**	**0.99**	**0.95**	**0.97**	**0.9767**
YelpCHI restaurant	(a) Bi-LSTM w/out optimization	95.93	0.99	0.92	0.95	0.9597
	(d) Bi-LSTM w/ SPSO (proposed GlOBiL)	**97.13**	**0.99**	**0.94**	**0.97**	**0.9714**

4.4. Experiments in Phase III: classification and testing

This phase labels the reviews as fake and real. The augmented and cleaned dataset includes an equal number of fake and real reviews. This dataset undergoes classification using the optimized Bi-LSTM classifier, evaluated through five-fold cross-validation. Optimal Bi-LSTM parameters, tailored for each dataset, are determined through SPSO by minimizing the validation error. Because the optimal values

Figure 6
Convergence graphs for PSO, HPSO, and SPSO across the four datasets



for parameters such as embedding dimensions, LSTM units, and dropout rates vary across datasets, this method ensures a customized model configuration for each dataset, enhancing classification performance. These optimal parameters for each dataset are given in Table 6 below. The performance on the test set is evaluated using metrics, including accuracy, precision, recall, F1 score, and ROC-AUC. The results of each fold are aggregated and summarized using the mean.

The proposed GlOBiL was compared with its baseline counterparts SMOTE–DNN, SMOTE–Bi-LSTM, BERT–Bi-LSTM, RoBERTa–Bi-LSTM, XLNet–Bi-LSTM, GPT-2–DNN, and GEMINI–Bi-LSTM. Further, GlOBiL was compared with previous research studies conducted on the four datasets. The next section presents the results of the above experiments.

5. Results and Discussion

This section evaluates the performance of the proposed framework, GlOBiL. Section 5.1 analyzes the performance and compares GlOBiL with its baseline counterparts in terms of precision, recall, F1 score, accuracy, and AUC across all four datasets. Section 5.2 compares the computational overhead by introducing SPSO in GlOBiL. Section 5.3 benchmarks the proposed framework against nine previous studies using the same metrics, demonstrating that it outperforms them all. Section 5.4 presents the ablation studies, and Section 5.5 concludes the results with parameter sensitivity analysis.

5.1. Comparison with baseline frameworks

GlOBiL's performance was analyzed across all four datasets: YelpZIP, YelpNYC, YelpCHI hotel, and YelpCHI restaurant. Table 7 shows the comparison of several baseline frameworks and the proposed GlOBiL framework on the task of detecting fake reviews using accuracy and AUC as performance metrics. The following observations can be made from Table 7.

1) The proposed framework is the best performer across all datasets in consideration, with the highest accuracy and AUC, thus providing the most accurate and reliable FRD.
2) Traditional frameworks using SMOTE struggle with complex language patterns, which can be seen in the lower accuracy and AUC. They have limitations in scalability and context understanding.

Table 6
Optimal parameters for each dataset

Dataset/parameter	Apply dropout	Embedding dimensions	LSTM units	Dropout rate
YelpZIP	0	63.14394814	232.1063245	0.250096849
YelpNYC	1	190.4019572	128	0.407878603
YelpCHI hotel	1	180.3076747	64	0.284268919
YelpCHI restaurant	1	139.0432249	85.0894018	0.166088246

Table 7
Accuracy and AUC scores of the proposed GlOBiL and various baseline frameworks across four datasets

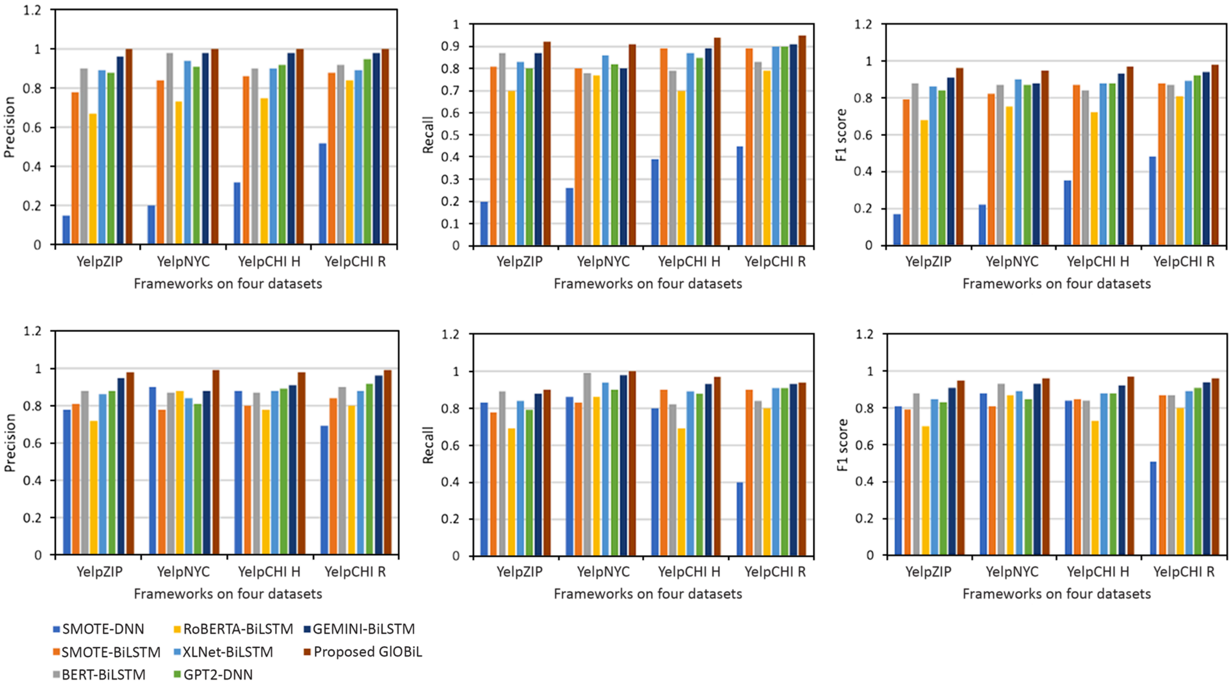
	YelpZIP		YelpNYC		YelpCHI hotel		YelpCHI restaurant	
Dataset/framework	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC
SMOTE–DNN	76.12	0.42132	79.00	0.48957	81.82	0.50372	85.32	0.61734
SMOTE–Bi-LSTM	79.43	0.75512	81.65	0.78167	81.99	0.79321	88.54	0.84003
BERT–Bi-LSTM	90.66	0.89272	91.70	0.92771	92.43	0.94720	92.33	0.95363
RoBERTa–Bi-LSTM	80.90	0.80218	83.39	0.81109	85.29	0.85384	87.90	0.88239
XLNet–Bi-LSTM	85.00	0.88329	89.15	0.90084	89.08	0.91849	89.09	0.92331
GPT-2–DNN	85.00	0.80328	86.00	0.81109	90.00	0.85903	90.00	0.90831
GEMINI–Bi-LSTM	92.70	0.93021	92.04	0.93152	94.11	0.95322	96.00	0.98384
Proposed GlOBiL	**95.37**	**0.95380**	**96.49**	**0.96480**	**97.67**	**0.97670**	**97.13**	**0.97140**

3) Transformer-based models such as BERT, RoBERTa, and XLNet, in combination with standard Bi-LSTM, understand bidirectional text, as is evident in the improved AUC scores.

The proposed framework is also evaluated based on precision, recall, and F1 score. Figure 7 displays the precision, recall, and F1 scores of the baseline frameworks in comparison to GlOBiL. It is observed that GlOBiL is the best framework across both fake and real classes, achieving the highest precision consistently across all datasets. GEMINI–Bi-LSTM is also a strong contender, with high precision in both classes, although slightly behind GlOBiL. BERT–Bi-LSTM also shows consistently high precision for both classes but is outperformed by the GEMINI and GlOBiL frameworks. SMOTE–DNN is the weakest model, with notably lower precision in both classes, especially in the fake class. Further, it can be seen that frameworks based on transformer architectures outperform traditional models such as SMOTE, demonstrating the effectiveness of advanced language generators in detecting fake reviews.

In addition, the bar charts compare the recall performance and show that GlOBiL has robust performance on both real and fake classes, particularly in YelpCHI datasets. SMOTE–DNN is another strong contender, delivering high recall on both real and fake classes, especially in YelpCHI restaurant. RoBERTa–Bi-LSTM and GPT-2–DNN show promising results as well, and XLNet–Bi-LSTM and GEMINI–Bi-LSTM demonstrate more variability and may not be as reliable across all datasets. F1 score is another metric that provides a balance between precision and recall. A higher F1 score indicates a good trade-off between identifying positive cases and avoiding false positives. The bar charts in Figure 7 representing the F1 scores show that GlOBiL is robust across all datasets, with the highest F1 scores in both classes. BERT–Bi-LSTM and RoBERTa–Bi-LSTM also perform

Figure 7
Precision, recall, and F1 scores (on fake and real classes) of GlOBiL and baseline frameworks across the four datasets



consistently well but are outperformed by GlOBiL. DNN frameworks tend to have lower F1 scores compared to Bi-LSTM-based frameworks, especially on the fake class. Predicting the real class appears to be easier for most frameworks, with more consistent and higher F1 scores compared to the fake class. This analysis shows the dominance of the proposed GlOBiL over its baseline counterparts.

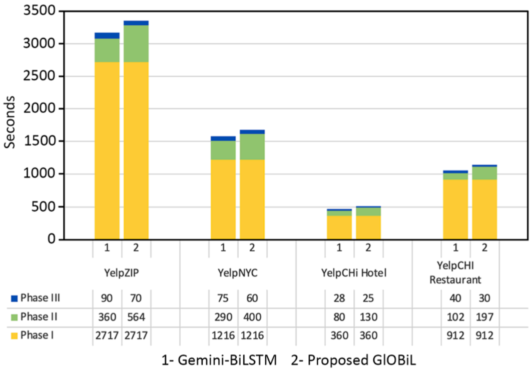
The superior performance of the proposed GlOBiL can be attributed to the integration of advanced language modeling, optimized classification, and class balancing strategies. Prior research has shown that transformer-based models such as GPT-2 possess strong capabilities in generating fluent and contextually coherent text [58], which makes them effective for synthetic data generation in class-imbalanced scenarios [59]. In addition, the Bi-LSTM classifier, used in GlOBiL, has been widely recognized for its ability to model sequential dependencies in text classification tasks [60]. The use of the proposed SPSO improves upon classical PSO and hybrid approaches by deferring binary updates, which as supported in delayed optimization literature [61], can reduce early-stage instability and improve convergence. These design choices align with the empirical improvement given in the following subsections.

5.2. Computational efficiency and performance trade-off

The proposed GlOBiL framework was designed to improve classification performance without increasing model complexity in terms of parameter count. The Bi-LSTM classifier used in Phase III retains the same number of trainable parameters as in any other baseline framework. SPSO performs hyperparameter optimization in Phase II across a mixed search space (binary and continuous) and does not modify the Bi-LSTM architecture itself. To assess the practical impact of this optimization, a comprehensive runtime analysis comparing GlOBiL to Gemini–Bi-LSTM was conducted. The reason for choosing Gemini–Bi-LSTM is that it achieved the highest classification accuracy across datasets among all models evaluated. Hence, it can be taken as the strongest comparator for runtime and performance evaluation. The runtime comparison results are shown in Figure 8.

Results show that the inclusion of SPSO resulted in a 5%–8% increase in total training time across all datasets. For example, on the YelpZIP dataset, the training time increased modestly from 3167 s (Gemini–Bi-LSTM) to 3351 s (GlOBiL) and similarly on YelpNYC from 1581 to 1676 s. Despite this modest overhead, GlOBiL delivered significant improvements in classification performance: accuracy gains of +2.67% on YelpZIP, +4.45% on YelpNYC, +3.56% on YelpCHI hotel, and +1.13% on YelpCHI restaurant (in Table 7) compared to Gemini–Bi-LSTM. In addition, GlOBiL achieved consistently higher F1 scores as shown in Figure 7. The performance gains of GlOBiL were even more pronounced than other baselines.

Figure 8
Component-wise runtime comparison of the proposed framework with Gemini–Bi-LSTM



Overall, although SPSO introduces additional training time, it does so without increasing the model complexity, and the modest computational cost is justified by the substantial improvements in classification accuracy and reliability across datasets.

5.3. Comparison with previous research works

This subsection presents a comparison of the proposed GlOBiL's performance against nine previous research methods on the metrics accuracy, precision, recall, F1 score, and AUC. Table 8 shows the results of this comparison on YelpZIP, YelpNYC, YelpCHI hotel, and YelpCHI restaurant datasets.

It is observed that the proposed framework GlOBiL demonstrates strong and consistent performance across all benchmark datasets compared to prior methods. For instance, on the YelpZIP dataset, it achieves the highest accuracy (95.37%) and AUC (0.95), significantly surpassing other methods such as DeepSpot [25], which has an AUC of 0.88, and FRDGMPM [28], which has an accuracy of 86%. Similarly, on the YelpNYC dataset, GlOBiL achieves an accuracy of 96.49% and an AUC of 0.96, outperforming WSEM-S [16] and FRDGMPM [28] in all metrics. In the YelpCHI hotel and restaurant datasets, GlOBiL maintains its superior performance with accuracies of 97.67% and 97.13%, respectively, and near-perfect precision and F1 scores. This consistent superiority across various datasets highlights GlOBiL's robustness and effectiveness in detecting fake reviews compared to traditional methods and recent frameworks.

In summary, the proposed GlOBiL framework consistently outperforms all other methods across multiple datasets, including YelpZIP, YelpNYC, YelpCHI hotel, and YelpCHI restaurant, in terms of accuracy, AUC, precision, recall, and F1 score. It demonstrates good ability to detect both fake and real reviews with high accuracy and minimal errors.

5.4. Ablation studies

This subsection analyzes the impact of four ablations in GlOBiL by systematically removing individual components to assess their contribution to the overall performance. The first ablation involves not optimizing Bi-LSTM as the classifier. The second focuses solely on optimizing the continuous parameters of Bi-LSTM, and the third ablation targets only the binary parameters of Bi-LSTM. Finally, the fourth ablation substitutes BERT embeddings in place of GloVe embeddings. Figure 9 illustrates the comparison between these ablations with the proposed framework across all four benchmark datasets.

Ablation 1 – standard Bi-LSTM (no optimization): this setup uses a standard Bi-LSTM classifier without optimization. The parameters are predefined with fixed values: dropout disabled (apply dropout = 0), embedding dimensions set to 50, LSTM units set to 128, and dropout rate of 0.1. These specific values have been chosen to reflect a commonly used baseline configuration in neural network experiments. Embedding dimensions of 50 are lightweight for efficient computation, and 128 LSTM units ensure sufficient model capacity. A dropout rate of 0.1 balances regularization without overpenalizing the network. This configuration serves as a benchmark to compare the impact of optimization.

Ablation 2 – optimization with continuous PSO only: here, dropout is disabled, and only the continuous parameters of Bi-LSTM are optimized using continuous PSO.

Ablation 3 – optimization with binary PSO only: this setup focuses exclusively on optimizing the binary parameters of Bi-LSTM

Table 8
Comparison of the proposed GlOBiL with prior research works on benchmark datasets

Benchmark dataset	Method	Accuracy (%)	AUC	Precision	Recall	F1 score
YelpZIP	DeepSpot (Nayak et al. [25])	-	-	0.86	0.88	0.86
	SVM w/ oversampling (Budhi et al. [62])	86	-	0.83	0.86	0.80
	BP (SVM) + undersampling (Budhi et al. [62])	79	-	-	-	-
	EMKL at Z = 20 (Zhang et al. [17])	-	0.84	0.83	0.81	0.82
	M-SMOTE with XGBoost (Kumar et al. [23])	-	0.83	0.77	0.74	0.78
	ImDetector (Zhang et al. [10])	-	0.88	0.85	0.83	0.84
	FRDGMPM (Luo et al. [28])	86	-	0.99	0.86	0.92
	Proposed GlOBiL	**95.37**	**0.95**	**0.99**	**0.91**	**0.95**
YelpNYC	SVM w/ oversampling (Budhi et al. [62])	89	-	0.84	0.89	0.84
	BP (SVM) w/ undersampling (Budhi et al. [62])	82	-	-	-	-
	WSEM-S (Singhal & Kashef [16])	95	-	0.90	**0.97**	0.92
	FRDGMPM (Luo et al. [28])	89	-	**1.00**	0.89	0.94
	SBiLM (Gupta et al. [24])	81	0.49	0.50	0.54	0.52
	Proposed GlOBiL	**96.49**	**0.96**	0.99	0.93	**0.95**
YelpCHI hotel	LR w/ oversampling (Budhi et al. [62])	86	-	0.81	0.86	0.80
	AB w/ undersampling (Budhi et al. [62])	77	-	-	-	-
	RUS-RF w/ opt param (Yao et al. [18])	-	-	-	-	0.71
	(RF + GBDT + lightGBM) (Yao et al. [18])	-	-	-	-	0.72
	FRDGMPM (Luo et al. [28])	86	-	0.99	0.86	0.92
	Proposed GlOBiL	**97.67**	**0.98**	**0.99**	**0.95**	**0.97**
YelpCHI restaurant	LR w/ oversampling (Budhi et al. [62])	86	-	0.82	0.86	0.81
	AB w/ undersampling (Budhi et al. [62])	85	-	-	-	-
	RUS-RF w/ opt param (Yao et al. [18])	-	-	-	-	0.78
	RF + lightGBM + CatBoost (Yao et al. [18])	-	-	-	-	0.79
	Proposed GlOBiL	**97.13**	**0.97**	**0.99**	**0.94**	**0.97**

using binary PSO. Continuous parameters remain fixed at their default values as given in ablation 1.

Ablation 4 – using a BERT embedding layer instead of GloVe: in this ablation, the GloVe embedding layer is replaced with the BERT embedding layer to evaluate the impact of different embedding techniques. The rest of the GlOBiL framework remains the same and continues to utilize the optimized Bi-LSTM.

The observations are as follows:

Ablation 1 (w/out optimizing Bi-LSTM) offers strong performance with balanced precision and recall but is outperformed by the proposed framework in almost all metrics. The reason for the good precision and recall is the inherent nature of Bi-LSTM of capturing long-range dependencies and patterns in text. However, it lacks optimization that could further boost its performance.

Ablations 2 and 3 also display less efficient performance because each approach neglects critical parameters. This partial optimization restricts the model's ability to explore and converge effectively in the search space, leading to suboptimal results.

Ablation 4 (using BERT embedding) shows a significant drop in performance even after optimizing Bi-LSTM, with lower precision, recall, accuracy, and AUC, indicating that it is less suitable for the task. The reason for this can be attributed to BERT's deep contextual embeddings, leading to overfitting or the introduction of noise. BERT embeddings may be less suited to distinguish between subtle differences between fake and real.

GlOBiL (the proposed framework) outperforms the others, achieving the highest scores in nearly every category, making it the best model for this classification task based on these metrics. It combines the strength of optimizing Bi-LSTM, which captures better semantic meaning and GloVe embeddings.

Each component of the proposed framework is necessary and contributes to the overall performance, robustness, and effectiveness of the system by addressing specific aspects of the optimization process.

5.5. Sensitivity analysis

This subsection discusses the impacts of four hyperparameters on GlOBiL's performance. To analyze sensitivity, a specific parameter is varied while keeping all other parameters fixed at their optimal values as determined by the proposed framework. The graphs in Figure 10 provide an analysis of the impact of the hyperparameters on GlOBiL across different datasets. The bar charts in the first row depict the effect of enabling or disabling a dropout layer. Enabling it generally improves performance, emphasizing its role in mitigating overfitting and enhancing generalization across various datasets.

The line charts in subsequent rows examine the effects of continuous hyperparameters such as embedding dimensions, LSTM

Figure 9
Comparison of the four ablation frameworks with GlOBiL

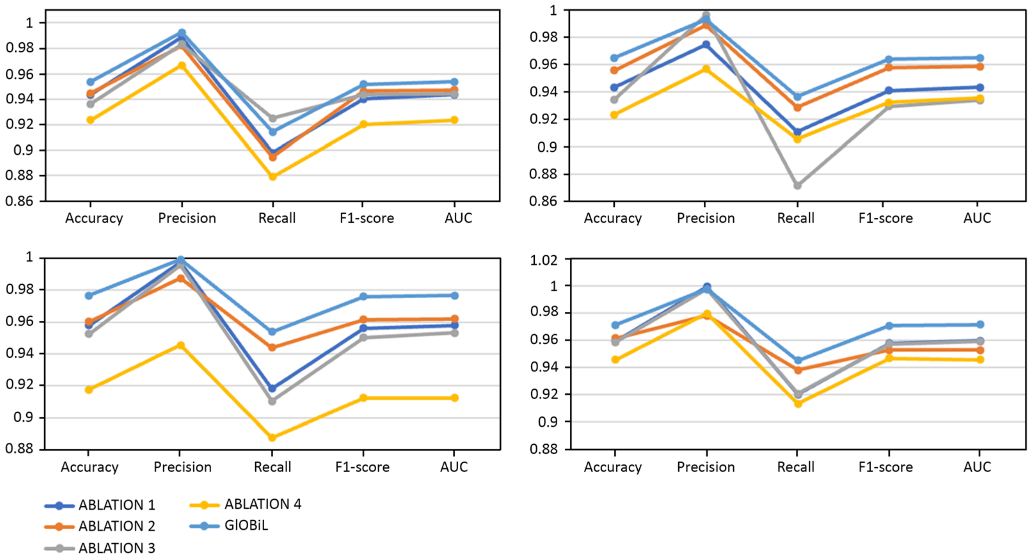
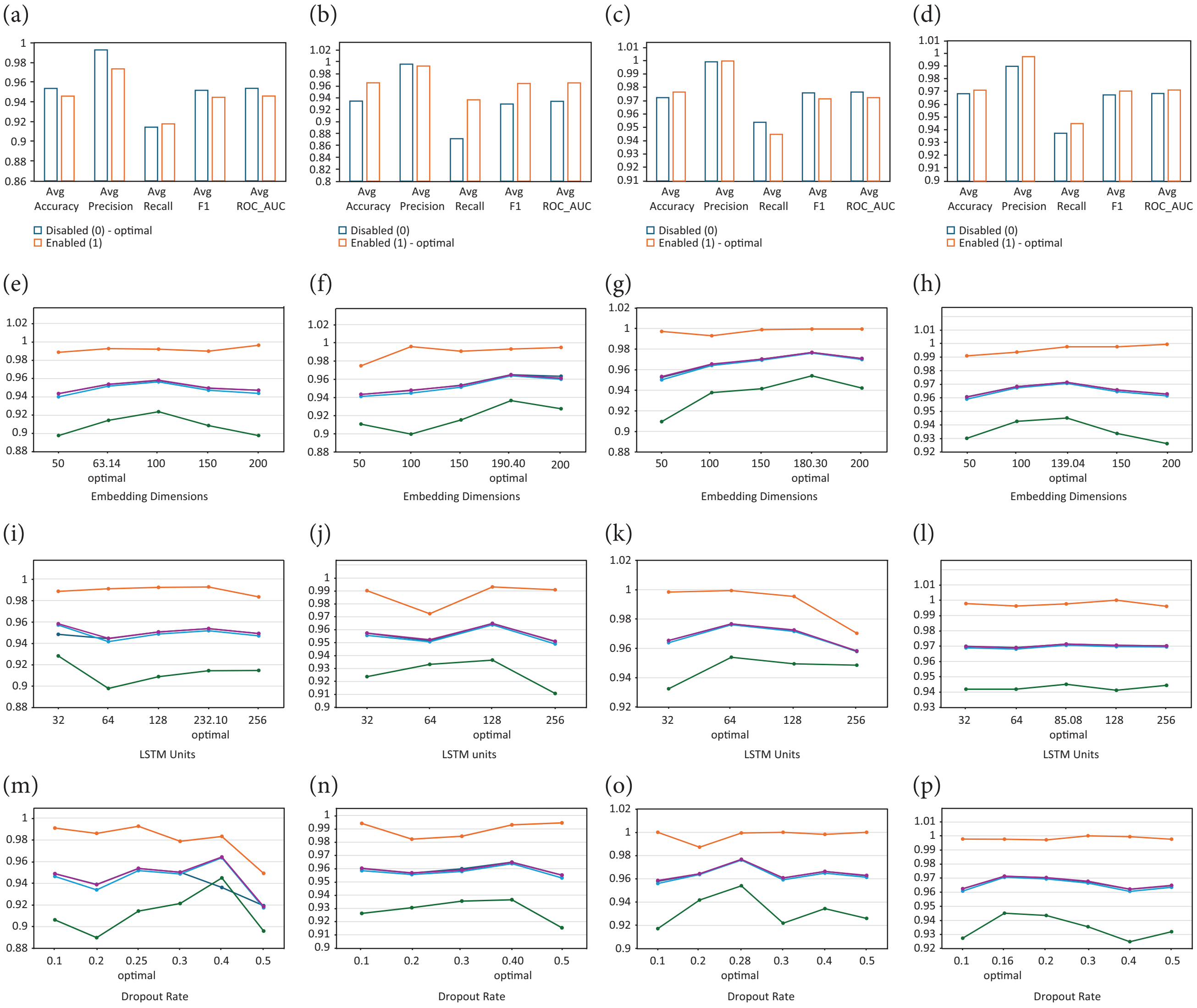


Figure 10
Sensitivity analysis of various parameters on the performance of GlOBiL



units, and dropout rates on performance metrics. Increasing embedding dimensions and LSTM units generally enhances performance metrics, but beyond a certain threshold, the improvement plateaus or declines, indicating diminishing returns. Dropout rates also exhibit a critical balance, where moderate dropout improves generalization but excessive dropout reduces model performance. These observations emphasize the importance of optimizing the Bi-LSTM parameters to maximize the proposed framework's performance for different datasets. The experimental results demonstrate that the proposed framework is effective in detecting fake reviews, showing improvements over baseline methods and previous research. Its consistently high precision and recall scores across multiple datasets indicate robustness and a good ability to generalize.

6. Conclusion and Future Work

Identifying fraudulent reviews is a significant challenge for stakeholders, including customers, online retail platforms, and businesses operating online, specifically in case of imbalanced datasets. This research introduces a novel framework called GlOBiL to handle the class imbalance problem in FRD. It utilizes a GPT-2 language generator with GloVe embedding and an optimized Bi-LSTM classifier. GlOBiL also incorporates the proposed novel SPSO algorithm to optimize the Bi-LSTM classifier, leading to faster convergence and improved classification accuracy. GlOBiL operates in three key phases. In the first phase, the original dataset is augmented with synthetically generated, contextually rich reviews to balance the distribution of real and fake reviews. The second phase uses GloVe embeddings to embed the augmented dataset and novel SPSO algorithm to optimize the Bi-LSTM classifier. The staggered approach of SPSO ensures better exploration of the search space, reducing parameter interference and improving convergence toward optimal solutions. In the final phase, the embedded, augmented dataset is used to train the optimized Bi-LSTM, which then classifies test reviews as fake or real.

The experiments were performed on several baseline methods. The results show that GlOBiL outperforms baseline methods and prior research across four benchmark imbalanced Yelp datasets. The results demonstrate that GlOBiL strikes an effective balance between precision and recall, resulting in a higher F1 score, which is crucial for maintaining the credibility and trustworthiness of online review platforms. The framework effectively maintains a low false positive rate, accurately identifying genuine reviews without misclassifying them as fake. It also demonstrates a low false negative rate, efficiently detecting fake reviews without confusing them with real ones. This highlights the framework's precision in minimizing misclassifications. In addition, GlOBiL achieves high accuracy and AUC scores across all four datasets, demonstrating its robustness and generalizability. Future research could extend this study by exploring additional datasets to further enhance the understanding of the FRD problem.

Although GlOBiL has demonstrated strong performance across four well-established labeled Yelp datasets, the availability of labeled, imbalanced datasets in other domains such as Amazon or TripAdvisor is limited. The existing datasets in these domains are either balanced or proprietary. Future research will focus on using semisupervised or transfer learning to evaluate the framework's generalizability across other platforms.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://doi.org/10.1145/2783258.2783370, reference number [8].

Author Contribution Statement

Richa Gupta: Conceptualization, Methodology, Software, Investigation, Resources, Writing – original draft. **Indu Kashyap:** Supervision, Writing – review & editing. **Vinita Jindal:** Resources, Writing – review & editing, Supervision.

References

[1] Kumar, A., & Saroj, K. (2020). Impact of customer review on social media marketing strategies. *International Journal of Research in Business Studies*, *5*(2), 105–114.

[2] He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, *41*(5), 896–921. https://doi.org/10.1287/mksc.2022.1353

[3] Shen, R. P., Liu, D., & Shen, H. S. (2023). Detecting review manipulation from behavior deviation: A deep learning approach. *Electronic Commerce Research and Applications*, *60*, 101283. https://doi.org/10.1016/j.elerap.2023.101283

[4] Zaman, M., Vo-Thanh, T., Nguyen, C. T., Hasan, R., Akter, S., Mariani, M., & Hikkerova, L. (2023). Motives for posting fake reviews: Evidence from a cross-cultural comparison. *Journal of Business Research*, *154*, 113359. https://doi.org/10.1016/j.jbusres.2022.113359

[5] Banerjee, S. (2022). Exaggeration in fake vs. authentic online reviews for luxury and budget hotels. *International Journal of Information Management*, *62*, 102416. https://doi.org/10.1016/j.ijinfomgt.2021.102416

[6] Gupta, R., Jindal, V., & Kashyap, I. (2024). Recent state-of-the-art of fake review detection: A comprehensive review. *The Knowledge Engineering Review*, *39*, e8. https://doi.org/10.1017/S0269888924000067

[7] Kotiyal, B., Pathak, H., & Singh, N. (2023). Debunking multi-lingual social media posts using deep learning. *International Journal of Information Technology*, *15*(5), 2569–2581. https://doi.org/10.1007/s41870-023-01288-6

[8] Rayana, S., & Akoglu, L. (2015). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 985–994. https://doi.org/10.1145/2783258.2783370

[9] Cardoso, E. F., Silva, R. M., & Almeida, T. A. (2018). Towards automatic filtering of fake reviews. *Neurocomputing*, *309*, 106–116. https://doi.org/10.1016/j.neucom.2018.04.074

[10] Zhang, W., Xie, R., Wang, Q., Yang, Y., & Li, J. (2022). A novel approach for fraudulent reviewer detection based on weighted topic modelling and nearest neighbors with asymmetric Kullback–Leibler divergence. *Decision Support Systems*, *157*, 113765. https://doi.org/10.1016/j.dss.2022.113765

[11] Shafie, M. R., Khosravi, H., Farhadpour, S., Das, S., & Ahmed, I. (2024). A cluster-based human resources analytics for predict-

ing employee turnover using optimized artificial neural networks and data augmentation. *Decision Analytics Journal*, *11*, 100461. https://doi.org/10.1016/j.dajour.2024.100461

[12] Akhavan, F., & Hassannayebi, E. (2024). A hybrid machine learning with process analytics for predicting customer experience in online insurance services industry. *Decision Analytics Journal*, *11*, 100452. https://doi.org/10.1016/j.dajour.2024.100452

[13] Khandokar, I. A., & Shatabda, S. (2023). New boosting approaches for improving cluster-based undersampling in problems with imbalanced data. *Decision Analytics Journal*, *8*, 100316. https://doi.org/10.1016/j.dajour.2023.100316

[14] Salminen, J., Kandpal, C., Kamel, A. M., Jung, S. G., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, *64*, 102771. https://doi.org/10.1016/j.jretconser.2021.102771

[15] Budhi, G. S., Chiong, R., Wang, Z., & Dhakal, S. (2021). Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. *Electronic Commerce Research and Applications*, *47*, 101048. https://doi.org/10.1016/j.elerap.2021.101048

[16] Singhal, R., & Kashef, R. (2023). A weighted stacking ensemble model with sampling for fake reviews detection. *IEEE Transactions on Computational Social Systems*, *11*(2), 2578–2594. https://doi.org/10.1109/TCSS.2023.3268548

[17] Zhang, W., Qin, G., & Wang, Q. (2021). Handling imbalance in fraudulent reviewer detection based on expectation maximization and KL divergence. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 421–427. https://doi.org/10.1145/3498851.3498989

[18] Yao, J., Zheng, Y., & Jiang, H. (2021). An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization. *IEEE Access*, *9*, 16914–16927. https://doi.org/10.1109/ACCESS.2021.3051174

[19] Cao, N., Ji, S., Chiu, D. K., & Gong, M. (2022). A deceptive reviews detection model: Separated training of multi-feature learning and classification. *Expert Systems with Applications*, *187*, 115977. https://doi.org/10.1016/j.eswa.2021.115977

[20] Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 497–501.

[21] Harris, C. G. (2012). Detecting deceptive opinion spam using human computation. In *HCOMP@AAAI*.

[22] Li, S., Zhong, G., Jin, Y., Wu, X., Zhu, P., & Wang, Z. (2022). A deceptive reviews detection method based on multidimensional feature construction and ensemble feature selection. *IEEE Transactions on Computational Social Systems*, *10*(1), 153–165. https://doi.org/10.1109/TCSS.2022.3144013

[23] Kumar, A., Gopal, R. D., Shankar, R., & Tan, K. H. (2022). Fraudulent review detection model focusing on emotional expressions and explicit aspects: Investigating the potential of feature engineering. *Decision Support Systems*, *155*, 113728. https://doi.org/10.1016/j.dss.2021.113728

[24] Gupta, R., Kashyap, I., & Jindal, V. (2024). SBiLM: Siamese Bi-LSTM model for handling imbalance in fake review detection. *Procedia Computer Science*, *235*, 1157–1166. https://doi.org/10.1016/j.procs.2024.04.110

[25] Nayak, A., Chen, H., Ruan, X., & Ouyang, J. (2019). DeepSpot: Understanding online opinion spam by text augmentation using sentiment encoder-decoder networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Analytics for Local Events and News*, 1–10. https://doi.org/10.1145/3356473.3365187

[26] Cheng, L. C., Wu, Y. T., Chao, C. T., & Wang, J. H. (2024). Detecting fake reviewers from the social context with a graph neural network method. *Decision Support Systems*, *179*, 114150. https://doi.org/10.1016/j.dss.2023.114150

[27] Xu, S., Cuan, H., Yin, Z., & Yin, C. (2024). A hybridized approach for enhanced fake review detection. *IEEE Transactions on Computational Social Systems*, *11*(6), 7448–7466. https://doi.org/10.1109/TCSS.2024.3411635

[28] Luo, J., Luo, J., Nan, G., & Li, D. (2023). Fake review detection system for online E-commerce platforms: A supervised general mixed probability approach. *Decision Support Systems*, *175*, 114045. https://doi.org/10.1016/j.dss.2023.114045

[29] Keya, A. J., Wadud, M. A. H., Mridha, M. F., Alatiyyah, M., & Hamid, M. A. (2022). AugFake-BERT: Handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification. *Applied Sciences*, *12*(17), 8398. https://doi.org/10.3390/app12178398

[30] Atliha, V., & Šešok, D. (2020). Text augmentation using BERT for image captioning. *Applied Sciences*, *10*(17), 5978. https://doi.org/10.3390/app10175978

[31] Sawai, R., Paik, I., & Kuwana, A. (2021). Sentence augmentation for language translation using GPT-2. *Electronics*, *10*(24), 3082. https://doi.org/10.3390/electronics10243082

[32] Cohen, S., Presil, D., Katz, O., Arbili, O., Messica, S., & Rokach, L. (2023). Enhancing social network hate detection using back translation and GPT-3 augmentations during training and test-time. *Information Fusion*, *99*, 101887. https://doi.org/10.1016/j.inffus.2023.101887

[33] Mulla, N., & Gharpure, P. (2023). Leveraging well-formedness and cognitive level classifiers for automatic question generation on Java technical passages using T5 transformer. *International Journal of Information Technology*, *15*(4), 1961–1973. https://doi.org/10.1007/s41870-023-01262-2

[34] Barnard, J. (2024). What are word embeddings? Retrieved from: https://www.ibm.com/think/topics/word-embeddings

[35] Pak, A., Ziyaden, A., Saparov, T., Akhmetov, I., & Gelbukh, A. (2024). Word embeddings: A comprehensive survey. *Computación y Sistemas*, *28*(4), 2005–2029. https://doi.org/10.13053/cys-28-4-5225

[36] Chen, Y., & Yin, B. (2025). Transformer-based fake news classification: Evaluation of DistilBERT with CNN-LSTM and GloVe embedding. *Informatica*, *49*(25). https://doi.org/10.31449/inf.v49i25.7710

[37] Ellaky, Z., Benabbou, F., Matrane, Y., & Qaqa, S. (2024). A hybrid deep learning architecture for social media bots detection based on BiGRU-LSTM and GloVe word embedding. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3430859

[38] Muka, G., & Mukala, P. (2024). Leveraging pretrained word embedding models for fake review identification. *Journal of Artificial Intelligence*, *6*, 211–223. https://doi.org/10.32604/jai.2024.049685

[39] Khan, F. Y., & Shaikh, T. A. (2024). Efficient and automated fake review detection with BERT embeddings. In *International Conference on Artificial Intelligence and Networking*, 179–191. https://doi.org/10.1007/978-981-96-2015-9_12

[40] Abduljaleel, I. Q., & Ali, I. H. (2025). Detecting fake news using BERT word embedding, attention mechanism, partition and overlapping text techniques. *TEM Journal*, *14*(2). https://doi.org/10.18421/tem142-16

[41] Ala'M, A. Z., Mora, A. M., & Faris, H. (2023). A multilingual spam reviews detection based on pre-trained word embedding and weighted swarm support vector machines. *IEEE Access*, *11*, 72250–72271. https://doi.org/10.1109/ACCESS.2023.3293641

[42] Deshai, N., & Bhaskara Rao, B. (2023). Unmasking deception: A CNN and adaptive PSO approach to detecting fake online reviews. *Soft Computing*, *27*(16), 11357–11378. https://doi.org/10.1007/s00500-023-08507-z

[43] Sirra, K. K., Mogalla, S., & Madhuri, K. B. (2024). CSSL-nO: Cat swarm sea lion optimization-based deep learning for fake news detection from social media. *International Journal of Information Technology*, *16*(7), 4225–4241. https://doi.org/10.1007/s41870-024-01943-6

[44] Saidi, R., Jarray, F., Kang, J., & Schwab, D. (2022). GPT-2 contextual data augmentation for word sense disambiguation. In *Pacific Asia Conference on Language, Information and Computation*.

[45] Bartoli, A., & Medvet, E. (2020). Exploring the potential of GPT-2 for generating fake reviews of research papers. In J. Antonio & Tallón-Ballesteros (Eds.), *Fuzzy systems and data mining VI* (pp. 390–396). IOS Press. https://doi.org/10.3233/FAIA200717

[46] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[47] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, *32*. https://doi.org/10.48550/arXiv.1906.08237

[48] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., …, & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv Preprint: 1907.11692*. https://doi.org/10.48550/arXiv.1907.11692

[49] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

[50] Pichai, S., & Hassabis, D. (2023). Introducing Gemini: Our largest and most capable AI model. Retrieved from: https://blog.google/technology/ai/google-gemini-ai/#performance

[51] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681. https://doi.org/10.1109/78.650093

[52] Yiqing, L., Xigang, Y., & Yongjian, L. (2007). An improved PSO algorithm for solving non-convex NLP/MINLP problems with equality constraints. *Computers & Chemical Engineering*, *31*(3), 153–162. https://doi.org/10.1016/j.compchemeng.2006.05.016

[53] Abdulqader, M., Namoun, A., & Alsaawy, Y. (2022). Fake online reviews: A unified detection model using deception theories. *IEEE Access*, *10*, 128622–128655. https://doi.org/10.1109/ACCESS.2022.3227631

[54] Mohawesh, R., Tran, S., Ollington, R., & Xu, S. (2021). Analysis of concept drift in fake reviews detection. *Expert Systems with Applications*, *169*, 114318. https://doi.org/10.1016/j.eswa.2020.114318

[55] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). What Yelp fake review filter might be doing?. In *Proceedings of the International AAAI Conference on Web and Social Media*, *7*(1), 409–418. https://doi.org/10.1609/icwsm.v7i1.14389

[56] Salminen, J., Mustak, M., Jung, S. G., Makkonen, H., & Jansen, B. J. (2025). Decoding deception in the online marketplace: Enhancing fake review detection with psycholinguistics and transformer models. Journal of Marketing Analytics, 1–18. https://doi.org/10.1057/s41270-025-00393-8

[57] Becerra-Rozas, M., Lemus-Romani, J., Cisternas-Caneo, F., Crawford, B., Soto, R., Astorga, G., …, & García, J. (2022). Approaches used in adapting metaheuristic optimization algorithms developed for continuous problems to discrete problems. *Mathematics*, 129.

[58] Bandyopadhyay, A., Chakraborty, P., Debdas, S., Patra, M., Mohapatra, S., & Guha, D. (2023). Beyond words: Harnessing GPT-2 to continue stories with imagination. In *2023 IEEE Silchar Subsection Conference (SILCON)*, 1–6. https://doi.org/10.1109/SILCON59133.2023.10404413

[59] Suhaeni, C., & Yong, H. S. (2023). Mitigating class imbalance in sentiment analysis through GPT-3-generated synthetic sentences. *Applied Sciences*, *13*(17), 9766. https://doi.org/10.3390/app13179766

[60] Kashid, S., Kumar, K., Saini, P., Dhiman, A., & Negi, A. (2022). Bi-RNN and Bi-LSTM based text classification for Amazon reviews. In *International Conference on Deep Learning, Artificial Intelligence and Robotics*, 62–72. https://doi.org/10.1007/978-3-031-30396-8_6

[61] Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle swarm optimization: A comprehensive survey. *IEEE Access*, *10*, 10031–10061. https://doi.org/10.1109/ACCESS.2022.3142859

[62] Budhi, G. S., Chiong, R., & Wang, Z. (2021). Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features. *Multimedia Tools and Applications*, *80*(9), 13079–13097. https://doi.org/10.1007/s11042-020-10299-5

How to Cite: Gupta, R., Kashyap, I., & Jindal, V. (2025). An Analytical Framework for Addressing Imbalance in Fake Review Detection Using Augmented Text and Swarm Optimized Classifier. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52026221