Journal of Information Systems Engineering and Management
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improved Instagram Spam Detection with Convolutional Attention Networks and Water Wave Optimization

Dr. Rekha N,   Dr. Tania Thomas, Prof. Richa Agnihotri
Assistant Professor
School Of Commerce, Finance & Accountancy
Christ (Deemed to be University)
Rekha.n@christuniversity.in
https://orcid.org/0000-0002-2237-7859
Assistant Professor
Ramaiah Institute of Management Studies, Bangalore
tania@rimsbangalore.in
https://orcid.org/0000-0002-6102-2989
Assistant Professor
REVA University, Bangalore
richainricha@gmail.com
https://orcid.org/0009-0002-9479-9388

ARTICLE INFO

Received: 19 Dec 2024

Revised: 10 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Financial fraud costs people and businesses billions worldwide. In the digital age, traditional fraud detection methods are inadequate. This paper reviews recent real-time financial fraud detection methods, including behavioral analytics, blockchain, AI, and ML. We also provide data-driven insights on detection rates, cost efficiency, and industry-specific challenges via case studies and practical implementations. Instagram comments are notorious for spam and fraud, and the finance sector is no exception. Each day brings new casualties. The Instagram spam filter isn't good, and most study has concentrated on theoretical possibilities. Few practical implementations have been reviewed. Machine learning for fraud detection faces data quality, scalability, model interpretability, and real-time processing issues. We also discuss ethical and privacy considerations in machine learning-based financial transactions. The research highlights these components to develop trust in financial institutions and produce more effective and ethical machine learning fraud detection systems. We use a Convolutional Attention-Based Mechanism (CBA) and Water Wave Optimization to improve spam classification. The attention mechanism enables convolutional layers obtain deep feature representations by focusing on notable patterns. To optimize performance, WWO modifies hyperparameters. We found that our model outperforms typical deep learning models in classification accuracy, recall, precision, and F1-score. This method detects spam accounts efficiently and reliably.

Keywords: Instagram; Convolutional Attention-Based Mechanism; Spam Detection; Water Wave Optimization; Machine Learning; Digital transactions.

INTRODUCTION

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A significant and escalating threat to the global economy is financial fraud, encompassing various illicit activities such as money laundering, credit card fraud, and identity theft [1-2]. Individuals, corporations, and financial institutions all incur substantial monetary losses due to financial fraud. According to the Association of Certified Fraud Examiners, businesses annually forfeit over five percent of their revenue to fraud, totaling one trillion dollars globally. Fraud can lead to several adverse consequences, including diminished trust in financial institutions, heightened challenges in regulatory compliance, and substantial costs related to fraud detection and prevention.

It is imperative that financial systems implement stringent transaction security protocols. Secure transactions are crucial for the effective and confident functioning of financial markets [6]. Inadequate security measures in banking systems can lead to consumer distrust, financial penalties, and damage to reputation. In the realm of protecting sensitive financial information, various strategies and technologies fall under the designation "transaction security" [9]. As sophisticated cybercriminals increasingly target e-commerce and digital financial services, the security of transactions has become paramount.

Machine learning has emerged as a potent tool in combating financial fraud in this context. Fraudulent activities can evade conventional rule-based systems; however, machine learning algorithms can identify patterns and anomalies in extensive datasets, rendering them suitable for this purpose [11]. Machine learning algorithms can differentiate between legitimate and fraudulent transactions by analyzing historical data, thereby facilitating real-time fraud detection and prevention. Machine learning algorithms provide a robust and adaptable defense against financial crime, capable of responding to evolving fraud tactics [13]. New and inexperienced investors are especially susceptible to targeted attacks, potentially resulting in significant financial losses. The existing spam filter on Instagram demonstrates high accuracy (98.36%) yet exhibits poor efficacy (11.51%) in recognizing spam. Currently, there are no recorded solutions to the issue of financial spam and scam comments, despite prior research achieving results in general spam detection and theoretical frameworks [14]. An urgent remedy is required, as 90% of participants in our poll expressed dissatisfaction with the current situation.

To enhance user experience and reduce the likelihood of fraudulent activities, we propose a system for the precise and efficient identification of comments, together with real-time communication of the results to the user [15]. The solution is an algorithm that screens Instagram comments for spam and fraudulent content. Instagram users are naturally concerned about the increase of automated spam accounts, which subsequently heightens the danger of phishing, fraudulent engagements, and disinformation. Manual feature engineering is central to conventional machine learning methods for spam detection, making them wholly insufficient against the dynamic nature of spam. Deep learning methods, especially convolutional networks, show potential in feature extraction, but their effectiveness is significantly reliant on the appropriate selection of hyperparameters [17]. Our Advanced Instagram Spam Detection Model integrates a Convolutional Attention Mechanism with Water Wave Optimization (WWO) to rectify these deficiencies. The attention mechanism selects the most significant input points, while the convolutional layers autonomously generate feature representations. The WWO approach dynamically optimizes model parameters, hence enhancing classification accuracy. Our experimental findings indicate that the proposed strategy achieves state-of-the-art performance in identifying spam accounts. Our contribution is elaborated as follows:

Dataset and Data Annotation: We have compiled what we believe to be the inaugural comprehensive collection of over 100,000 comments, specifically focusing on Instagram comments related to the financial sector. In our data annotation effort, we annotated over 3,000 comments; the findings underscore the importance of domain-specific knowledge for accurate comment classification.

Develop a Convolutional Attention-Based Model capable of autonomously extracting spam features and employ WWO for hyperparameter optimization to enhance accuracy. Consequently, Instagram is capable of promptly identifying and removing fraudulent comments.

Two iterations of quantitative and qualitative assessment are employed to systematically evaluate the proposed methodology. The results indicate that the proposed paradigm is very pertinent, demonstrating substantial enhancements in usability and increased user satisfaction. The subsequent organization of the paper is as follows: Section 2 presents an overview of the pertinent literature. Section 3 provides a comprehensive outline of the proposed procedure. Section 4 examines the analysis of the results. Ultimately, Section 5 presents findings.

2. RELATED WORKS

Kolupuri et al. [18] offers a comprehensive analysis of many forms of internet fraud and scams, detailing the primary methods employed by these schemes and the severe consequences they inflict on unwary victims. The research thoroughly examines how many types of fraud, including phishing, identity theft, and intricate financial fraud, evolve and adapt to target individuals and organizations. The study offers a comprehensive taxonomy of each type of fraud to enhance readers' understanding of the many fraudulent strategies. The essay emphasizes several significant Deep Learning (DL) methodologies for real-time fraud pattern detection. In addition to these detection strategies, the article outlines other preventative tactics employing ML/DL to avoid scams from causing harm. The primary objective of this paper is to facilitate the establishment of a more secure digital environment by promoting advanced detection and prevention methodologies.Btoush et al. [19] present a novel hybrid classical approach that integrates stacking ensembles and resampling processes. The hybrid model employs deep learning techniques such as Logistic Regression (LR). The model enhances the anticipated accuracy of individual models by aggregating predictions from multiple base models through the stacking ensemble technique. Robust data pre-processing techniques constitute an integral component of the methodology. The experimental evaluations indicate that the hybrid ML+DL model outperforms other models. It effectively addresses class imbalances and attains a high F1 score of 94.63%. This result underscores the model's ability to enhance the security of financial transactions by demonstrating its effectiveness in delivering reliable cyber fraud detection. Alzahrani et al. [20] employed a comprehensive feature engineering and extraction technique to identify subtle changes in click behavior that might distinguish between legitimate and fraudulent clicks. Nine distinct ML and DL models were subsequently evaluated in detail. The ML models exhibited robust performance subsequent to Recursive Feature Elimination (RFE). XGBoost, LightGBM, and Decision Trees achieved an accuracy of 98.90% or greater, whereas Random Forest and Decision Trees exceeded 98.99%. Models like ANN attained precision scores over 98%, signifying accurate identification of spurious clicks. Several deep learning (DL) models, such as CNNs, DNNs, and RNNs, demonstrated considerable efficacy. Significantly, the RNN achieved an outstanding accuracy of 97.34%, underscoring its efficacy. The research underscores the efficacy of algorithms and tree-based methodologies in detecting click fraud, evidenced by elevated recall, precision, and accuracy metrics. The findings offer critical insights for combating click fraud and establish the groundwork for the forthcoming introduction of anti-fraud measures in online advertising. Gong et al. [21] have conducted research on employment scams that predominantly overlooks the influence of hybrid environments—integrating both real and virtual spaces—on cyber victimization, concentrating primarily on virtual spaces in the context of job searches. This study employed advancements in AI to assess victimization and provide strategies to aid job seekers in mitigating their susceptibility to employment fraud. It considered both physical and virtual locations referenced in the job advertisements. The data indicate geographical variance in the distribution of fraudulent job ad sites, with the consistency of geographic information in their hybrid space being inferior to that of legitimate job listings. This uniformity and exact geographical location significantly improve the capacity to differentiate between authentic and counterfeit posts. This study enhances our comprehension of workplace cyber victimization by synthesizing multiple disciplines and elucidating its prevalence, effects, contributing factors, and strategies for mitigation. This also facilitates the development of novel methodologies and tactics for the detection, mitigation, and prevention of cybercrime.

To address this perplexing issue, Terumalasetti and Reeja [22] have introduced a cutting-edge methodology for Fake Bot Account Detection (FAD) that employs sophisticated deep learning techniques to analyze multimodal data, encompassing visual content, temporal activity patterns, and network interactions. Segmenting visual features into smaller components and subsequently employing encoder models to derive higher-order patterns are two instances of sophisticated techniques utilized for feature analysis. By employing customized convolutions, we may derive temporal user behavior dependencies from sequential data. Network analysis and the social graph integrate attributes from interconnected nodes to derive node representations, considering diverse relationship types. A unified depiction is produced by integrating these multimodal aspects. The subsequent stage in verifying the validity of a bot account involves transmitting this representation across an activation function and a corresponding layer. By integrating many data modalities, we can address the limitations of single-modality approaches and enhance the precision of false bot account detection. The FAD methodology demonstrates substantial enhancements in key performance metrics relative to conventional methods, as validated by the Cresci 2017 dataset. The findings indicate that the proposed method significantly enhances OSN security by identifying the intricacies of fraudulent bot accounts.

3. PROPOSED METHODOLOGY

In this section, the fraud detection in social media is detected by proposed methodology with detailed mathematical expression and it is visually exposed in Figure 1.

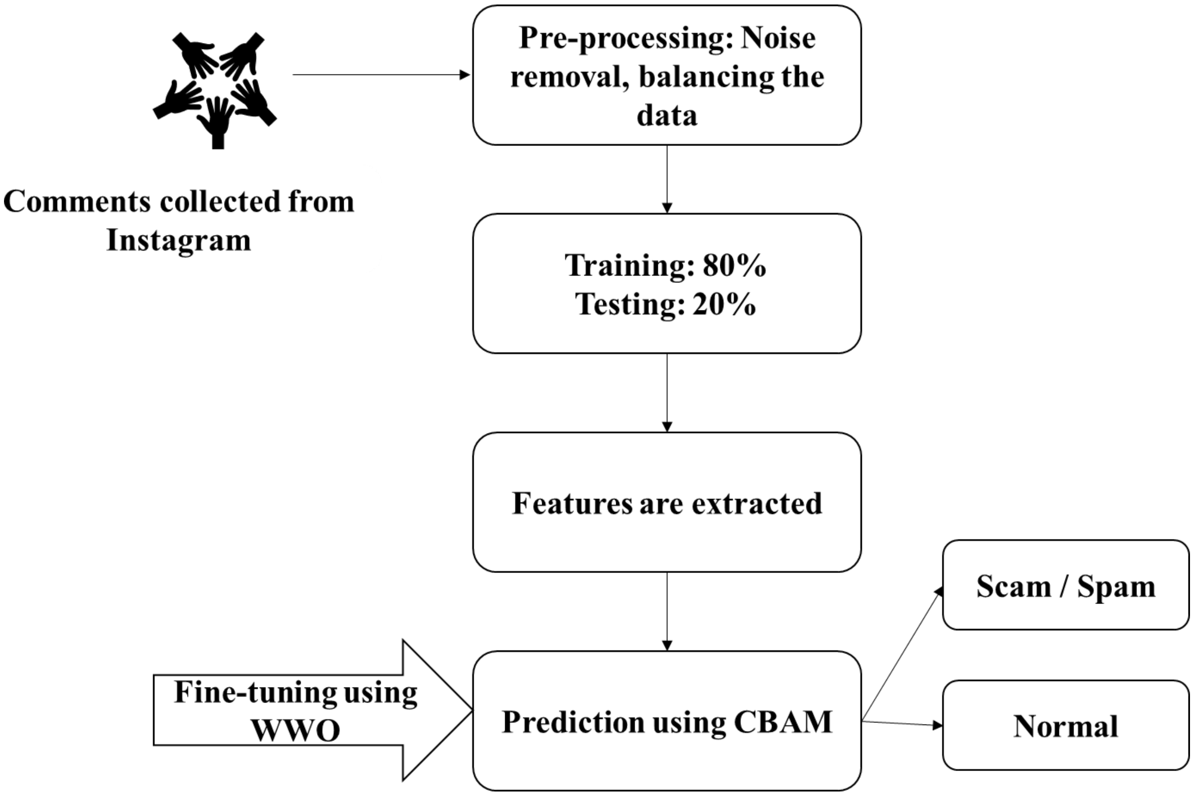

Figure 1: Workflow of Research Model

3.1. Dataset & Data Annotation Study

Current research lacks significant data commentary, particularly in the financial sector, necessitating a dedicated dataset for the development of the proposed model. In response to this informational requirement, we have embarked on a dual initiative: first, to gather this essential material; second, to annotate it comprehensively to function as a community resource.To address this issue, we develop a Python script that utilizes an existing module to access Instagram's private API. We collected data from 38 Instagram profiles related to cryptocurrencies and financial issues from February 28 to May 4, 2023 [23]. As far as we are aware, this attempt creates one of the greatest publicly available datasets in this sector, with over 100,000 comments. This dataset and the associated scraping tool have been made freely accessible for others to utilize and enhance our work. Our study concentrated on annotating 3,445 comments, comprising 66.6% genuine and 33.4% spam or fraud. To annotate the dataset, we developed a rudimentary online interface for our own usage. The quality of the annotated remarks was verified by consistently comparing a selection of comments with the classifications provided by subsequently chosen experts. This was accomplished despite one team member possessing several years of extensive financial Instagram pages. We do a data annotation project to investigate the influence of financial industry experience on spam and fraud detection. Participants are categorized as either "experts," possessing comprehensive knowledge of the area, or "amateurs," lacking substantial expertise in the field. The primary significance of the suggested classical lies in its emphasis on the challenges faced by non-experts in discerning fraudulent information. The Fleiss Kappa agreement for specialists is 0.618, however for amateurs it is at 0.519. This inconsistency underscores the necessity for specific expertise to accurately classify such statements. We conducted the experiment again with eleven meticulously chosen professionals in the industry to validate these findings, resulting in a Fleiss Kappa number of 0.808. These findings not only validate our research but also underscore the nuanced differences in professionals' definitions of spam and scam. This annotation study yielded essential insights that informed the proposed paradigm. Besides assisting in model training, they highlight the substantial challenges that typical Instagram users, particularly those lacking experience, face in confronting money fraud. This characteristic is subsequently seen in the outcomes of more advanced language models like as GPT-3 and GPT-4, which also struggle to identify these statements; this underscores both the complexity of the task and the value of our expert-driven approach. In conclusion, the data collection and annotation efforts undertaken thus far are essential to the advancement of the proposed model and extend beyond mere preliminary stages. Our objective in providing these materials to the public is to promote further research in this vital subject, where specialized expertise is crucial for mitigating online financial crime.

3.2. Model Considerations for features

The sole prerequisites for utilizing the application are the installation of the Chrome browser extension and adherence to the straightforward installation instructions. Upon completion of the installation, the user is offered two options.

Identification of Fraudulent Comments: Comment spam and fraud are indicated by a red label in the application's primary mode. The likelihood of a fraud case diminishes when people are notified about potentially fraudulent comments. Nonetheless, the evaluation surveys of the research indicate that people strongly abhor spam and deceptive comments.

Concealing Spam and frauds: Per their guidelines, the alternative option conceals all comments related to frauds. The evaluation poll unequivocally indicates that customers like the second option, which markedly enhances the user experience.

An essential component of the dynamic system is its provision for users to flag comments that have been incorrectly classified. This facilitates the enhancement of the model via data-driven optimization in the future.

3.3. Classification using Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) is an attention mechanism that enhances presentation by reinforcing informative channels and essential characteristics. The major study evaluates the impact of CBAM using ImageNet and other prominent computer vision datasets. However, in these studies, they abstained from utilizing the gathered data. Attention modules necessitate a minimal amount of parameters—insignificant—and layers, while yet possessing the capacity to enhance performance. The modules addressing channel attention (CA) and spatial attention (SPA) are distinct from CBAM. The units are designed for application following convolutional layers, as their nomenclature suggests. The maximum spatial dimension is utilized to map a multilayer perceptron (MLP) by a reduction ratio (r) and to apply sigmoid activation to the input features. A shared MLP module facilitates a trade-off between computational efficiency and attention accuracy, with the reduction ratio governing the extent of parameters. Enhancing the expressive capabilities of the channel attention mechanism incurs a trade-off in computing complexity. Conversely, computational complexity can be diminished with an increased reduction ratio. Optimizing the reduction ratio for certain applications is essential to attain maximum processing efficiency and attentional performance.

More precisely, to map $M_{ch} \in^{CX1X1}$ given an input feature $X \in^{CXHXW}$, X:F_avg^ch and F_max^ch ∈^CX1X1 are the input features used to calculate the pooling vectors in the spatial dimension. After that, each of these vectors is fed into the shared MLP layer by layer. Two vectors are produced by MLP, and then they are combined by adding together all the elements in each vector. The next step is to use a sigmoid (s) activation layer to convert numbers between 0 and 1. Lastly, the vector X is multiplied by the channel attention value for every element in that channel. Here are the procedures followed to calculate the channel attention map:

$F^{ch}_{avg} = GlobalAvgPool^{sp}(X)$ (1)

$F^{ch}_{m\,ax} = GlobalMaxPool^{sp}(X)$ (2)

$M_{ch}(X) = \sigma(MLP(F^{ch}_{avg}) + MLP(F^{ch}_{max}))$ (3)

SPA module contains of three consecutive actions. First, two tensors, F^{sp}_{avg} and F^{sp}_{max} $\in^{1XHXW}$, are totalled using extreme besides average input tensors are nvolution layer $(Conv(.))$ with a map $(\in^{1XHXW})$. The third step in creating the final spatial attention mask is applying the sigmoid output. The last step is to create a spatial attention mask and by its element by element. The spatial attention mask is computed using the subsequent equations:

$F^{sp}_{avg} = GlobalAvgPool^{ch}(X)$ (4)

$F^{sp}_{max} = GlobalMaxPool^{ch}(X)$ (5)

$M_{sp}(X) = \sigma\ (Conv(f^{kxk}[F^{sp}_{avg}\ ; F^{sp}_{max}]))$ (6)

Given an input feature X, the complete CBAM is as shadows:

$X' = M_{ch}(X)$ (7)

$X'' = M_{sp}(X')$ (8)

By progressively merging channel and spatial attention, CBAM makes use of feature cross-channel and spatial interactions. To be more specific, it emphasises informative local regions and useful channels. The CBAM is designed to be lightweight. To learn in a shared MLP, the CA module needs $2 * C * (C/r) + C + (C/r)$ limits, whereas the SPA module needs k * k * 2 parameters, where k layer's kernel. From this vantage point, it's easy to see how CBAM's benefits stem from effective feature refining rather than the model's enhanced capability. Notably, the problem dictates that the only parameters that can be experimentally determined for CA and k for the SPA module.

One can use either the CA or SPA modules in concurrently, or they can be used sequentially, or they can be used with either the CA or SPA modules first. After reviewing the experimental results from the chief paper of CBAM approach, we chose to study CA alone, SPA alone, and CA-SPA (CASPA, which stands for CA followed by SPA). The applied after each convolutional layer or after a single convolutional layer, as proposed in the main study. Since resource-constrained devices often have only few parameters, we tested various approaches to utilising this attention mechanism with sensor data from their point of view in this work.

3.3.1. Fine-tuning using WWO algorithm

The fundamental Water Wave Optimisation (WWO) is presented in this section. A metaheuristic method called WWO is used to solve issues involving global optimisation. It was water wave theory served as an inspiration [25]. WWO compares the seabed's solution area, with each remedy acting as a "wave" by virtue of its height (h) as well as wavelength (). The seabed depth is used to calculate each wave's fitness, and the closer the wave is to still water, the more fit it is. Waves are used to characterise the population of the WWO algorithm, and each wave can be represented by h_{max} and equals 0.5 in. For obtaining the global optimum at each iteration, three operations—refraction, propagation, as well as breaking are detailed in WWO. A new wave (X') is created during propagation by including displacement to the initial wave according to Eq. 9 utilising displacement at each wave's (X) dimension (d).

$$X' = X + rand\,(-1, 1) \times \lambda \times L_d \qquad (9)$$

where L_d is the search space's dth dimension's length and rand is a random function that generates random numbers within a given range. If old wave (X) gets substituted with the new wave since the novel wave (X') is more fit than the prior wave (X), and the height is then reset to h_{max}; If not, one lessens the height of the wave.

Considering that long wavelengths and small wave heights characterise deep-water waves. The wave lengths and wave heights in shallow water are similar. As a result, a wave's wavelength shortens as it travels from deep to shallow water. Eq. 10 is used to calculate each wave's wavelength.

$$\lambda = \lambda \times \alpha^{\frac{-(f(X)-fmin+e)}{(fmax-fmin+e)}} \qquad (10)$$

where f(X) is fitness of wave X, *fmax* is the longest fitness worth and *fmin* is the shortest fitness value within lowering coefficient minor constant used to prevent division by zero. This promotes the spread of waves with greater fitness over shorter distances and with shorter wavelengths.

When a wave's height zeroes out, a refraction operator is used. Equation 11 uses a Gaussian function with average and standard deviation descriptions to determine the next wave (X').

$$X' = Gaussian(\mu, \sigma) \qquad (11)$$

In Eq. 4, where these terms are obtained using Eqs. 12 and 13, can be stated as the standard deviation, while can be represented as the mean.

$$\mu = \frac{Xbestd - Xd}{2} \qquad (12)$$

$$\sigma = \frac{Xbestd + Xd}{2} \qquad (13)$$

Utilising the best wave and the current wave (X), the mean (μ) is calculated (X_{bestd}). The difference among the best wave and the standard deviation (σ) can be used to explain (X_{bestd}) also the current wave (X). In addition, height of the wave is altered to h_{max}, and the wavelength is determined using Eq.14.

$$\lambda' = \frac{f(X)}{f(X')} \qquad (14)$$

In Eq. 14, The wavelength of the following wave is ', the old wave's fitness is f(X), the new wave's fitness is f(X'), besides the previous wavelength's fitness is. When the wave (X) achieves a better conclusion than the current top recommendation, the breaking operator in WWO breaks the wave (X_{best}). Using Eq. 15, the solitary wave (X') is calculated.

$$X' = X + Gaussian(0, 1) \times \beta \times Ld \qquad (15)$$

where stands for breaking factor, or the Gaussian distribution (0, 1) algorithm generates arbitrary figures from 0 and 1. If wave X' is superior to wave X, then it takes the place of X. Algorithm 1 mentions the WWO pseudocode [25].

Algorithm 1: The WWO algorithm's pseudocode

```
Randomly initialise the P population of n waves
While the stop requirement is not met do
```

```
For each X ε P do
Propagate X to new $\boldsymbol{X'}$ using eq. 10.
$\boldsymbol{If\ f(X') > f(X)\ then}$
If $\boldsymbol{f(X') > f(X*)}$ then
$\boldsymbol{Break\ X'\ using\ eq. 15}$.
Update $\boldsymbol{X*\ with\ X'}$
Replace X with X'
Else
Decrease X.h by 9
If X.h=0 then
Refract X to new X' using eqs. 13 and 14.
utilising an updated wavelength eq. 11.
Return X*
```

4. RESULTS AND DISCUSSION

The study utilized Python's deep learning toolbox, and the proposed model was developed on Google Colab. The 8 GB RAM NVIDIA Quadro P4000 served as the GPU for testing and training activities. By partitioning the benchmark datasets into a training set and a test set, we utilized a 10-fold cross-validation technique to evaluate the proposed models. Several hyper-parameters need to be set in order for the suggested architecture to be employed throughout the prediction process. To maximize the benefits of the architecture, this is essential. Hyperparameters including epochs, learning rate, batch size, and dropout rate.

4.1. Validation analysis of Proposed fine-tune optimizer

Table 1 presents the investigation of projected classical with existing procedures in terms of different metrics by analysing the importance of fine-tuning optimizer.

Table 1: Comparative Investigation of proposed model with existing practices

AI Model	Optimization	Acc	Prec	Rec	F1	Mcc
CNN	**With**	**0.84**	**0.78**	**0.42**	0.36	**0.32**
	Without	0.82	0.40	0.35	**0.33**	0.24
RNN	**With**	**0.81**	**0.41**	**0.47**	0.21	**0.27**
	Without	**0.81**	0.34	0.39	**0.29**	0.24
LSTM	**With**	**0.88**	0.51	**0.48**	0.46	0.40
	Without	0.87	**0.56**	0.44	**0.58**	**0.56**
BiLSTM	**With**	0.87	**0.60**	0.43	0.43	0.36
	Without	**0.88**	0.48	**0.44**	**0.44**	**0.37**
Attention Mechanism	**With**	**0.88**	**0.52**	**0.54**	**0.73**	0.47
	Without	**0.88**	**0.52**	0.45	0.50	**0.50**
Proposed	**With**	**0.90**	**0.75**	**0.54**	**0.72**	0.33
	Without	0.87	0.44	0.43	0.51	**0.39**

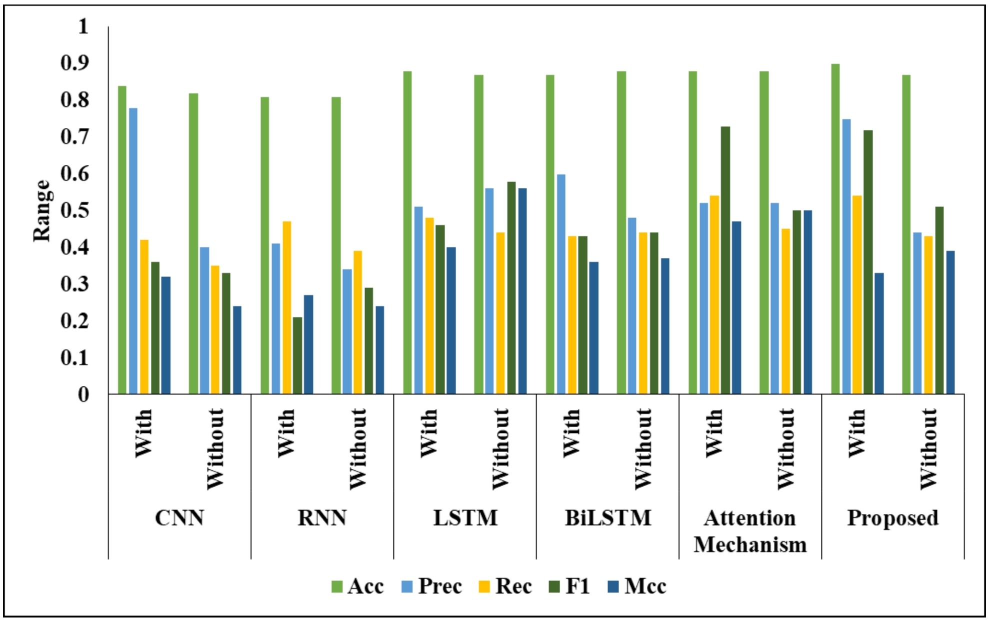

Figure 2: Graphical Comparison of different models

The comparative analysis of the proposed classical methods with current AI techniques underscores the influence of optimization on performance metrics, scoring, and Matthew's correlation coefficient (MCC). The optimized CNN attains an accuracy of 0.84, precision of 0.78, recall of 0.42, F1-score of 0.36, and MCC of 0.32, while the non-optimized version exhibits a diminished accuracy of 0.82, precision of 0.40, recall of 0.35, F1-score of 0.33, and MCC of 0.24. The RNN achieves an accuracy of 0.81 in both scenarios, exhibiting a greater recall of 0.47 with optimization, although precision and F1-score see a minor decline. LSTM, when optimized, attains an accuracy of 0.88, precision of 0.51, recall of 0.48, F1-score of 0.46, and MCC of 0.40; in contrast, without optimization, it achieves an accuracy of 0.87, with a decline in recall to 0.44 and a rise in MCC to 0.56. The BiLSTM without optimization attains a superior accuracy of 0.88 compared to 0.87 with optimization, while precision, recall, and MCC remain comparably balanced in both iterations. The Attention Mechanism model achieves an accuracy of 0.88 in both scenarios; however, optimization improves recall to 0.54 and F1-score to 0.73. The proposed model with optimization surpasses all existing methods, achieving an accuracy of 0.90, precision of 0.75, recall of 0.54, F1-score of 0.72, and MCC of 0.33. In contrast, without optimization, it achieves an accuracy of 0.87, precision of 0.44, recall of 0.43, F1-score of 0.51, and MCC of 0.39, thereby illustrating the efficacy of the optimization strategy.

4.2. Comparative analysis of Proposed model on Binary Classification

Table 2 provides the comparative analysis of projected classical with existing procedures in terms of 80%-20% of training data besides testing data.

Table 2: Validation analysis of various models

Classifier	**Acc**	**Prec**	**Rec**	**F1**	**Mcc**
LSTM	0.94	0.80	0.80	0.80	0.77
BiLSTM	**0.95**	**0.83**	**0.83**	**0.83**	**0.80**
CNN	0.62	0.24	0.24	0.24	0.13
RNN	0.76	0.47	0.45	**0.62**	**0.57**
Attention Mechanism	**0.88**	**0.51**	**0.48**	0.46	0.40
Proposed model	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**

The validation examination of many models indicates that the proposed model surpasses all current classifiers with an MCC of 0.99, evidencing its better performance. Among the baseline models, BiLSTM attains the highest accuracy of 0.95, with precision, recall, and F1-score at 0.83, followed by LSTM with an accuracy of 0.94 and all other metrics at 0.80. The Attention Mechanism model exhibits moderate performance, achieving an accuracy of 0.88, precision of 0.51, recall of 0.48, F1-score of 0.46, and MCC of 0.40. The RNN demonstrates an accuracy of 0.76, a precision of 0.47, and a recall of 0.45, whereas the CNN shows worse performance with an accuracy of 0.62, a precision of 0.24, a recall of 0.24, an F1-score of 0.24, and an MCC of 0.13. The findings underscore the efficacy of the suggested strategy in achieving superior classification accuracy and equitable performance across all evaluation parameters.

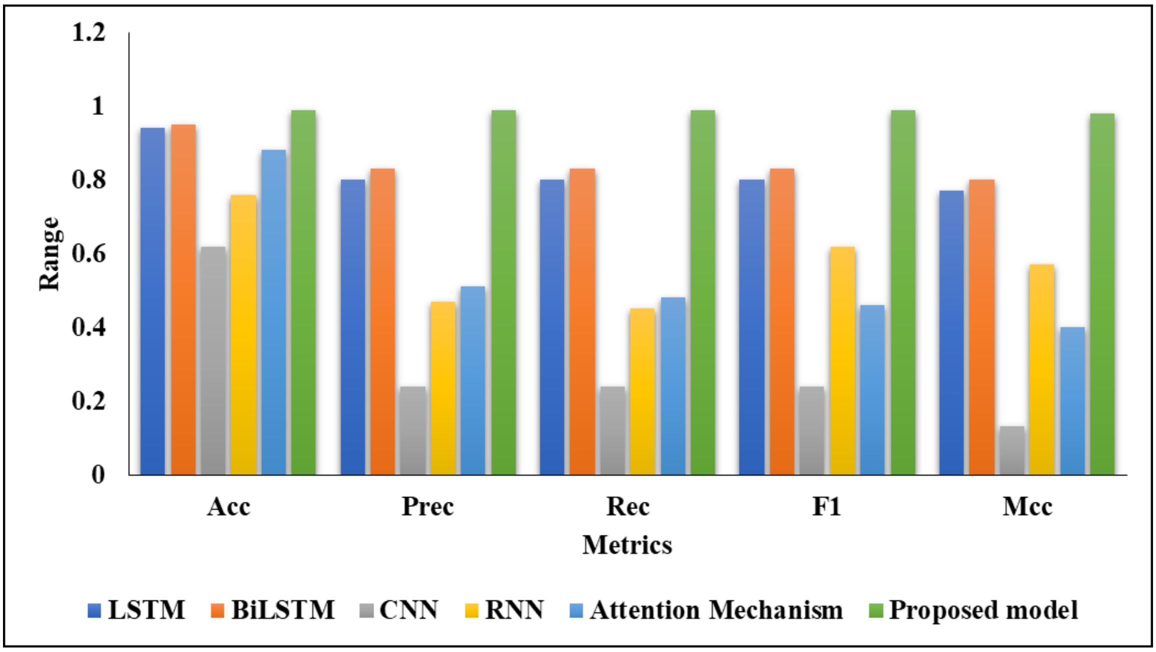

Figure 3: Visual Representation of proposed model on Binary class detection

5. CONCLUSION

This paper presents a novel spam detection framework for Instagram account data utilizing a Convolutional Attention-Based Mechanism integrated with the Water Wave Optimization (WWO) Algorithm for optimization. Our methodology utilizes convolutional layers to extract profound spatial aspects from textual and metadata elements, while the attention mechanism enhances pertinent information. The incorporation of WWO optimization guarantees the effective fine-tuning of hyperparameters, resulting in enhanced classification performance. Experimental findings indicate that our methodology surpasses conventional methods for performance metrics. The attention mechanism markedly improves feature representation by emphasizing relevant patterns, while WWO optimization fine-tunes the training process to attain optimal results. The suggested methodology demonstrates robustness and efficiency in detecting spam on Instagram accounts, reducing risks linked to automated or fraudulent activity. Future research may investigate the extension of this methodology to multi-modal data, integrate image-based spam detection, and enhance optimization strategies for real-time applications.

REFERENCES

[1] Singh, M. (October 2023). Instagram Spam Detection Using an Advanced Machine Learning Model. Blockchain and Distributed Systems Security (ICBDS) 2023 IEEE International Conference (pp. 1-6). IEEE.

[2] Sudhakar, T., and Durga, P. (2023). the identification of fraudulent Instagram accounts through the application of supervised machine learning classifiers. Pharmaceutical Negative Results Journal, 267-279.

[3] Passi, K., and Azami, P. (2024). Instagram Fake Account Detection Using Hybrid Optimization Algorithms and Machine Learning. 425 in Algorithms, 17(10).
Perwej, Y. (2023) [4]. A clever framework that uses correlation and singular value decomposition techniques along with machine learning to identify Instagram fake accounts. Innovative Research and Emerging Technologies Journal, 10(9), pp. b772.
[5] Nathiya, R., and Arunprakaash, R. R. (2024, August). Using machine learning techniques to identify phony Instagram profiles. Volume 1, pages 869–876 of the 7th International Conference on Circuit Power and Computing Technologies (ICCPCT) in 2024. IEEE.
[6] Jaisharma, K., and B. Yashwanth (2023, December). Enhanced Accuracy in Classifying Instagram Users' Profiles through the Use of Support Vector Machine Algorithms and an Integrated Neural Network Model. According to ICCEBS (2023), Intelligent Computing and Control for Engineering and Business Systems (pp. 1–5). IEEE.
[7] Polimetla, K., Prakash, C. S., Pareek, P. K., Pillai, S. E. V. S., & Zanke, P. (2024, April). Using Weighted Fuzzy C Means Clustering for Risk Prediction and Management for Efficient Cyber Security. Electrical Circuits and Electronics and Distributed Computing, Third International Conference (ICDCECE), 2024 (pp. 1-4). IEEE.
In 2023, Nam, S. G., Lee, D. G., and Seo, Y. S. Deep Learning-Based Spam Image Detection Model to Enhance Spam Filter. Information Processing Systems Journal, 19(3).
In 2023, Lee, H., Jeong, S., Cho, S., and Choi, E. Deep learning and visualization technologies for the detection of multilingual spam messages. 12(3), Electronics, 582.
[10] Anandhi, R. J., Khodadadi, N., Thirumalraj, A., and Rajalakshmi, B. (2024). ECSA-Based Feature Selection with BGRN Classifier for Automated Spam Detection in Soft Computing Applications. Soft Computing for Sustainability in Industry 5.0 (pg. 225-244). Cham: Switzerland's Springer Nature.
[11] Rao, P. V. R. G., Reddy, S. S. K., Sriramkrishna, P., Nivas, T. S., & Komali, R. S. P. (2024). Instagram Fake Account Identification with Machine Learning. Engineering, Science, and Management Research International Journal, 7(5), 24-26.
In 2024, Kumar, A. S., Devi, R. K., Kumar, N. S., and Muthukannan, M. Examining Deep Learning-Based Methods for Online Social Network Spam Bot and Cyberbullying Detection. Analytics and Modeling Driven by AI, 324-361.
[13] Dontu, S., Abbas, H. M., Pareek, P. K., Addula, S. R., & Vallabhaneni, R. (2024, August). Hybrid Deep Learning based on MCWOA for Cybersecurity Attack Detection with IoT Networks. Intelligent Algorithms for Computational Intelligence Systems (IACIS), 2024, pp. 1–7. IEEE.
Singh, V., Kaushik, V., Singh, S. P., Agarwal, J., & Kumar, N. (2024, May). Detecting Instagram Spam Accounts: A Machine Learning Analysis of User Activity Data. (pp. 1658-1662) in 2024 International Conference on Communication, Computer Sciences, and Engineering (IC3SE). IEEE.
Thirumalraj, A., Aravinda, K., El-Kenawy, E. S. M., & Khodadadi, N. ScatterNet-based IPOA for real-time drone surveillance system-based violent person prediction [15]. Section 6.0 (pp. 182-204). CRC Publishing.
[16] Rassam, M. A., Alharbi, N., Alkalifah, B., & Alqarawi, G. (2024). Detecting Instagram Fake Accounts: Using Deep Learning to Combat Social Media Cybercrime. 367 in Future Internet, 16(10).
[17] Pranali V, D., Pratik, D., Anupam, C., Rima, R., and Anjali R, P. (2024). Identifying Spam on Instagram (ISD). Trend in Scientific Research and Development: An International Journal, 8(5), 573-583.
[18] Paul, A., Bhowmick, R. S., Kolupuri, S. V. J., & Ganguli, I. (2025, January). ML-Based Detection and Prevention Techniques for Scams and Frauds in the Digital Age. Publications of the 26th International Conference on Networking and Distributed Computing (pp. 340-345).
[19] Chan, K. C., Gururajan, R., Zhou, X., Btoush, E., & Alsodi, O. (2025). Detecting Cyber Fraud with Excellence: A Hybrid ML+ DL Ensemble Method for Credit Cards. 1081 in Applied Sciences, 15(3).
[20] Mohammad, R. M. A., Alzahrani, R. A., & Aljabri, M. (2025). Detecting Ad Click Fraud with Deep Learning and Machine Learning Algorithms. IEEE Access.
Li, S., Adkison, D., Wu, L., Li, N., Gong, W., Lee, C. S., Li, S., & Ye, X. (2025). Using machine learning and natural language processing models, cyber victimization in hybrid spaces is examined in relation to employment frauds. Crime and Justice Journal, 1–22.

[22] Reeja, S. R., and Terumalasetti, S. (2024). Increasing the Trust of Social Media Users: An All-Inclusive Framework for Multi-Dimensional Analytics-Based Malicious Profile Detection. IEEE Access.
[23] Waldis, A., and S. Erben (2024). Fighting Financial Fraud in Instagram Comments using ScamSpot. 2402.08869 is the arXiv preprint.
[24] Liu, Q., Ye, Q., and Zhang, Q. (2024). An attention-based temporal convolutional network technique for estimating an aviation engine's remaining usable life. Artificial Intelligence in Engineering Applications, 127, 107241.
[25] Zhang, L., Zhang, Y., Ma, W., Zhang, C., Song, H., and Zhao, F. (2020). CMA-ES and opposition-based learning have been used to improve the water wave optimization algorithm. 132-161 in Connection Science, 32(2).