SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I8P134

Volume 12 Issue 8, 395-406, August 2025

Original Article

Moth Flame Optimization and CNN-GBM Fusion for Advanced Business Intelligence in Banking

Madhu Kumar Reddy P[1], M.N.V Kiranbabu[2]

[1,2]*Department of Computer Science & Applications, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.*

[1]*Corresponding Author : madhukumarreddy@gmail.com*

Received: 18 June 2025 Revised: 19 July 2025 Accepted: 20 August 2025 Published: 30 August 2025

Abstract - *In the era of digital banking, deriving actionable insights from customer complaints is essential to improve service quality, identify problems, and inform decision-making. Conventional methods are ineffective in handling high-dimensional and unstructured data, and their performance in real-world settings is compromised. This research proposes a hybrid deep learning model combining Moth Flame Optimization (MFO) and a CNN-GBM fusion model to overcome these limitations and deliver solid business intelligence from unstructured banking data. Preprocessing was performed on unstructured text data, mainly customer feedback, using common NLP methods such as stopwords removal, tokenization, and lemmatization. The pre-processed text was transformed into contextualized vector representations using BERT embeddings. Moth Flame Optimization was utilized to choose the most descriptive features, which minimized dimensionality and optimized the model's efficiency. A 1D Convolutional Neural Network (CNN) was utilized to obtain high-level features, which were trained using a Gradient Boosting Machine (GBM). The performance was measured with respect to baseline models like SVM and Random Forest. The new framework performed better, with the CNN+GBM model reaching an accuracy of 93.7% and an F1-score of 92.1%. MFO cut down the feature space by 67.5%, highly reducing training time while maintaining the accuracy of models. The value of ROC-AUC at 96.1% validated the high discriminatory capability. Comparative studies proved the new model outperforms traditional classifiers. This work illustrates how to leverage a combination of feature selection through MFO and the fusion of CNN-GBMs for unstructured banking data and business intelligence. Semantic richness comes through the adoption of BERT embeddings, complemented by computationally efficient assistance from MFO. The technique presents an interpretable and scalable banking solution available for institutions requiring inferences of information from bulky amounts of client-authored text.*

Keywords - *BERT embeddings, Business Intelligence, CNN-GBM, Moth Flame, MFO, ReLU.*

1. Introduction

In today's digital age, companies generate and collect massive amounts of information from a range of touchpoints, especially where customers are dealt with industries like finance, banking and e-commerce [1]. Much of this data comes in unstructured text forms like support requests, reviews on the web, complaint tickets and feedback surveys. Traditional BI systems have never been fully geared up to cope with such sources of data, despite offering valuable insights on the behavior of customers, the levels of satisfaction, inefficiencies at operations and new issues relating to services. These old-fashioned approaches often rely on rule-based analysis, structured data formats or elementary statistical summaries, none of which are adequate to generate deeper meaning from complex language-based inputs. Consequently, many companies miss out on significant opportunities to react to visionary client input that could influence strategic decisions. To cross this hurdle, interest is growing to integrate deep learning and advanced Natural Language Processing (NLP) techniques into Business Intelligence (BI) platforms so that unstructured data can be utilized to the fullest [2]. Organizations wanting to remain competitive in an increasingly data-driven world need to undergo this transformation.

With varying levels of success, much research has explored a variety of machine learning and natural language processing methods for text data analysis. For sentiment classification and topic detection applications, early research focused on traditional models such as logistic regression, Support Vector Machines (SVM), and decision trees. Although useful, these models often fail to adequately capture language's semantic and contextual richness. The capacity to capture word meaning in context has considerably risen with newer developments, notably the emergence of transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers). High-dimensional feature spaces, which may be computationally expensive and difficult to interpret, are brought about by the application of BERT embeddings [3]. By identifying the most relevant features and reducing noise within the input vectors, optimization strategies such as MFO present an effective means for addressing these limitations [4]. Also, Convolutional Neural Networks (CNNs) have proven to extract higher-level representations as well as neighbourhood patterns from sequence data; however, they are seldom optimized when combined with BERT and metaheuristic strategies. Secondly, even though they are known to be robust and precise, ensemble learning models such as Gradient Boosting Machines (GBM) remain underutilised in hybrid deep learning architectures for use in textual business intelligence. Consequently, an intensive plan integrating the above-discussed methods remains largely unexplored and untested in actual business domains such as banking.

The present work proposes a hybrid deep architecture involving CNNs for deep pattern learning, GBM for resilient classification, Moth Flame Optimization for feature selection and BERT for contextual text embeddings to bridge these research and implementation gaps. The objective is to design a scalable and efficient architecture that can support huge volumes of customer feedback data and generate reliable, interpretable outputs that enable data-driven decision-making in banking operations. Specifically, the research aims to utilize CNNs to detect local semantic trends in selected features, improve high-dimensional BERT embeddings via MFO for better computational efficiency and model interpretability, and utilize GBM to perform ultimate classification for predictive tasks such as sentiment classification, fraud risk detection or customer churn prediction. In order to illustrate the advantages in terms of precision, efficiency, and scalability, the proposed model will also be contrasted with traditional BI tools. Ultimately, this research improves the development of intelligent, next-generation BI systems that apply metaheuristic optimization and deep learning to derive actionable insights from unstructured sources of data within the financial sector. The key Contributions of the research are as follows,

- Proposed a novel deep learning framework integrating BERT, Moth Flame Optimization, CNN, and GBM to evaluate customer feedback.
- Utilized Moth Flame Optimization in selecting significant features from BERT embeddings and avoiding unnecessary information.
- Enhanced accuracy in predictions with the application of CNN for feature extraction and GBM for better decision-making.
- Developed an effective system utilizing large banking datasets, allowing business entities to attain valuable insights from customer complaint and review reports.

The paper is organized as follows: Section.II as Literature Review, Section.III as the Problem Statement, Section.IV as Methodology, Section.V as Result and Discussion, and Section. VI as a Conclusion.

2. Literature Review

Doush et al. [5] propose two new MFO-based optimization algorithms to optimize Multi-Layer Perceptron (MLP) training. The proposed algorithms solve problems such as slow convergence and local optima encountered in gradient descent. The techniques are used in the prediction of iron ore prices using feature reduction methods such as Pearson's and categorized correlation. The results demonstrate that the proposed MFO algorithms perform better compared to conventional swarm intelligence and classical machine learning algorithms. This methodology illustrates better MLP performance on demanding prediction problems.

Tang et al. [6] suggested that epilepsy diagnosis from EEG signals has been significantly advanced by machine learning models, although choosing the best hyperparameters is still problematic. Echo State Networks (ESNs) are good despite their sensitivity to the initialization of reservoir weights. A new work improves ESNs with MFO based on a new Feature Distribution Evaluation Function (FDEF). The latter assesses the quality of features according to class separability instead of classifier accuracy. The method presented attains higher accuracy and sensitivity on the Bonn and CHB-MIT EEG datasets.

Kumar et al. [7] provide a step-by-step approach to creating supervised learning models—Decision Tree, Random Forest, and K-Nearest Neighbor—to detect high-risk loan customers. False positives were minimized to prevent Non-Performing Assets (NPA) in the banking industry. Visualizations were also used to assist banks in understanding customer behavior and possibly forecasting client loyalty. As financial growth increases in developing nations, the demand for personal loans has been on the rise, and therefore, predicting loan default is vital for banks. Growing instances of fraud, as reported by the Reserve Bank of India, underscore the necessity of strong prediction systems.

Tunowski [8] suggests that Business Intelligence (BI) systems were analyzed regarding their contribution towards improving the commercial banks' financial sustainability. It was hypothesized in the research that the introduction of BI enhances the financial standing of the banks. A new comparative approach was used, which compared financial ratios over time, with industry norms, and within the framework of overall economic fluctuation. A synthetic indicator (ABI) was constructed from data from six large Warsaw Stock Exchange banks, accounting for 60% of the commercial bank assets in Poland. The results reaffirmed that BI positively affects productivity, asset and liability quality, profitability, and debt management of commercial banks.

Sathupadi et al. [9] Internet banking is increasing rapidly, making it necessary to have secure and efficient decision-making systems. This paper introduces BankNet, which applies big data technology and a BiLSTM neural network for analyzing transactions effectively. BankNet performs impressively well with an insignificant error rate and accurate fraud detection, maintaining good data speeds. BankNet assists banks in detecting fraud in a timely manner and enhances their operations. With its scalability and superior performance, BankNet enables more secure and credible online banking.

Li et al. [10] say that Credit risk forecasting of listed firms is now one of the foremost issues with the growing pace of financial markets. Basic models such as Z-score, Logit, and neural networks tend to show poor predictions. Recent work introduces deep models like CNN and LSTM for greater prediction accuracy regarding credit risk forecasting. Incorporation of an attention mechanism with the combination of CNN-LSTM enhances predictive performance further through optimizing information handling. Research indicates that this hybrid methodology performs better than conventional models and provides improved forecasts in credit risk prediction.

Liu et al. [11] introduce that churn rates rise in sectors such as finance, and acquiring a new user costs more than keeping an existing one. Artificial intelligence, especially deep learning, has been implemented to improve customer churn prediction. The paper presents a Bidirectional Long Short-Term Memory Convolutional Neural Network (BiLSTM-CNN) model combining Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), eliminating the drawbacks of standalone RNN and CNN models. The research also compares the performance of the BiLSTM-CNN model and that of an attention-based version (AttnBiLSTM-CNN) based on bank data. Experimental results indicate the superiority of the AttnBiLSTM-CNN model to the BiLSTM-CNN in terms of accuracy, churn rate, F1 score, and AUC, indicating the potential of the attention mechanism to enhance user churn prediction.

Cavus, Mohammed, and Yakubu [12] Suggests that Mobile banking applications are crucial for convenience but are characterized by low uptake in developing nations such as Nigeria. This research employs AI and Structural Equation Modeling (SEM) to determine the core determinants influencing mobile banking usage, drawing data from 823 participants. It establishes that risk, trust, facilitation conditions, and poor digital legislation are significant hindrances, while social influence and service quality have no effect. AI-informed models are superior to conventional models in describing these determinants. The findings offer implications for the enhancement of mobile banking adoption and security in Nigeria, with recommendations for future studies.

Oyarhossein et al. [13] introduce open banking as an open model under which banking information is exchanged using APIs between independent parties to boost market services. Although APIs have existed for a long time to serve financial applications, their use has primarily focused on information sharing and not on transferring funds. Advantages of open banking are increased customer experience, new revenue streams, and enhanced services for the underserved markets. Nevertheless, difficulties intervene when digital platforms such as Alibaba and WeChat continue to improve, with scarcity in data exchange working against progress. Overcoming the data exchange limitations is indispensable to open banking delivering its fullest value.

3. Problem Statement

In contemporary banking, the explosion of unstructured customer opinion in the form of online reviews and digital media is both an opportunity and a challenge [14]. Conventional classification techniques tend to fail to accurately capture subtle language, contextual sense, and implied customer issues in these reviews [15]. In addition, high-dimensional representations such as those produced by BERT, though rich in context, bring computational complexity and redundancy, degrading model efficiency and interpretability. Current machine learning classifiers, when utilized alone, fail to capture both high-level decision boundaries and deep semantic patterns [16]. Hence, the need for an optimized hybrid framework that can selectively reduce feature dimensions, identify deep textual patterns, and provide solid classification performance is of utmost importance. This study fills that gap by suggesting a new fusion method combining MFO for feature selection, 1D Convolutional Neural Network (CNN) for hierarchical pattern extraction, and Gradient Boosting Machine (GBM) for ultimate prediction. The objective is to improve business intelligence capacity in banking by facilitating accurate sentiment classification and customer insight extraction from raw textual reviews.

4. Materials and Methods

The approach utilizes a hybrid deep learning architecture to derive actionable information from unstructured customer feedback in banking. The process starts with gathering text data, such as customer reviews and ratings. The preprocessing steps, like normalizing, tokenizing, lemmatizing, and utilize pre-trained BERT embeddings to transform the text into contextual numerical vectors. For improved model performance, Dimensionality Reduction and feature selection are applied using MFO. These are fed to the feature-selected inputs via a 1D Convolutional Neural Network (CNN) to detect semantic patterns and produce high-level features. Then, the processed features are classified through a Gradient Boosting Machine (GBM) model, and classification performance is calculated through conventional metrics for classification. This approach seeks to produce worthwhile business intelligence insights such as customer opinion, churn forecasts, and market segmentation techniques. Figure 1 shows the Methodology Framework.

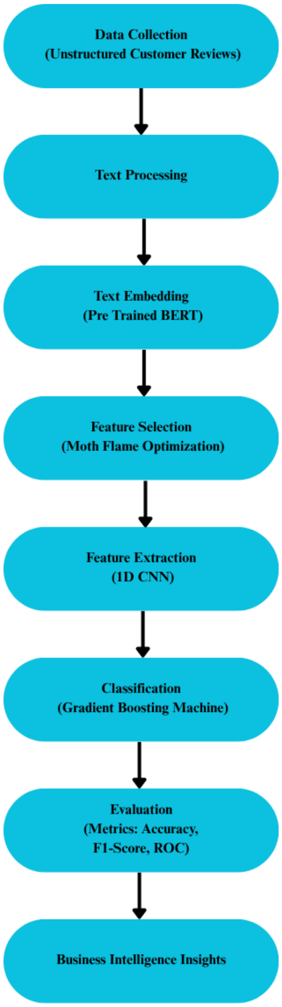

Fig. 1 Methodology framework

4.1. Data Collection

The Banks Customer Reviews Dataset, which is collected for the aim of customer sentiment analysis within the banking industry, is the primary data source for the current research. The "Banks Reviews Customer Dataset" [17] boasts a vast collection of over 1000+ data points of user-generated reviews and ratings spanning various banks. The dataset is made up of real user reviews based on personal experiences. The data set is predominantly unstructured text information, and it serves a critical function in the assessment of customer satisfaction, identification of operational loopholes, and enhancing the quality of services through advanced business intelligence techniques.

4.2. Text Preprocessing

In a bid to transform unstructured, raw customer opinions into an analyzable format, text preprocessing is vital. Cleaning and normalizing the data is very important to achieving effective downstream processing since user-created information is noisy and informal. The purpose of this phase is to reduce textual variability without losing the semantic meaning needed to achieve accurate feature extraction and classification.

4.2.1. Lower Casings

In order to maintain consistency across the corpus, all text is lowered to lowercase at the beginning of the preprocessing process.

4.2.2. Punctuations

The second step is to remove punctuation, numerical characters, and regular English stopwords such as "is," "and," and "the," which usually do not contribute anything useful to sentiment analysis.

4.2.3. Tokenization

The groundwork for lexical and semantic examination is then established by using tokenization to split every review into individual words or tokens. In embedding models, these tokens are the fundamental units of representation, and it is represented as,

$$\boldsymbol{Tokenized\,Text} = \boldsymbol{Tokenizer}(\boldsymbol{text}) \tag{1}$$

Where, $\boldsymbol{Tokenizer}$ is a function that splits the text into subwords or words based on a pre-defined vocabulary.

4.2.4. Lemmatization

Lemmatization is employed to further enhance the quality of the text. By converting words into their dictionary or base word (e.g., "running" to "run"), the technique facilitates broader generalization during feature learning and aids in clumping similar phrases. However, lemmatization is desired here because it is linguistically accurate, particularly for sentiment-rich data.

The unstructured consumer reviews are mapped into a uniform input form that can be infused with pre-trained models such as BERT and feature selection when these preparation steps are executed systematically.

4.2.5. Text Representation with BERT Embeddings

Classic word embeddings like TF-IDF and GloVe are typically applied to represent text in machine learning applications. The limitations of these approaches in representing the complex and dynamic relationships between words in a sentence have led to their shortcomings. TF-IDF, for example, operates on independent words without any context about their surroundings, while GloVe gives fixed vector representations to words from global co-occurrence statistics but remains context-insensitive. To mitigate these weaknesses, this work exploits BERT (Bidirectional Encoder Representations from Transformers), a cutting-edge technique that is intended to capture the contextual meaning of words in relation to the words surrounding them in the sentence.

BERT is a transformer model that reads text bidirectionally, i.e., it considers both the left and right context of a word at the same time, as opposed to earlier models like GPT, which consider left-to-right or right-to-left context. By employing a pre-trained BERT-base-uncased model, every preprocessed review is converted into a high-dimensional vector, usually of 768 dimensions. In this vector, the context-dependent meaning of each word is dynamically tuned depending on its context in the overall sentence, allowing the model to understand intricate relationships like polysemy (multiple meanings of a word) and contextual meaning shifts. With a context-aware approach, BERT can understand sentiment, intent, and other semantic properties of text much more accurately than earlier word embedding techniques. Figure 2 shows the BERT Framework.

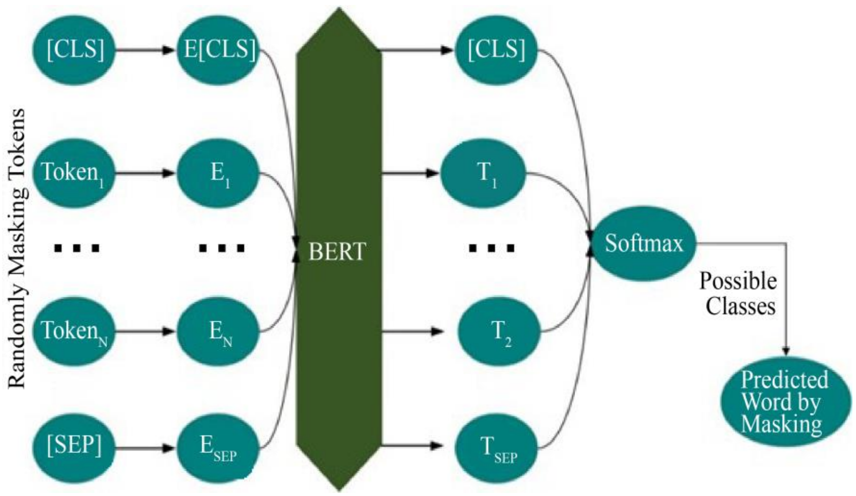

Fig. 2 BERT framework

Input Representation

BERT uses token embeddings, segment embeddings, and position embeddings to represent the input sequence. The input representation for any token $E_i$within a sequence is given as,

$$E_i = T_i + S_i + P_i \tag{2}$$

Where, T_i is the token embedding for the ith token, S_i is the segment embedding, P_i is the position embedding.

Self-Attention Mechanism

BERT uses a multi-head self-attention mechanism to compute the relationships between tokens in the input sequence. The self-attention score for a token i with respect to token j can be calculated as,

$$Attention(i,j) = softmax\frac{Q_i.K_j^T}{\sqrt{d_k}} \tag{3}$$

Final Embedding Representation

Once processed through several layers of attention and feedforward networks within the BERT model, the final representation of each token is derived by collecting information across all layers. The final embedding for a token can be represented as,

$$h_i = \ BERTE_i \tag{4}$$

Where, h_i is the final hidden state for the token.

Contextualized Embedding

Each token of the sequence outputs a dense vector of high dimensions that encodes the word's syntactic and semantic meaning within its context. This would be represented as,

$$v_i = h_i \tag{5}$$

Where, $\boldsymbol{v_i}$ is the context-aware representation.

Feature Selection Using MFO

The intrinsic redundancy in high-dimensional BERT embeddings compromises interpretability as well as computation, so it becomes important to select features. MFO and other biologically inspired optimization algorithms provide an attractive solution in which MFO cleverly traverses and capitalizes on the large subset feature space. The "moth" subsets are led by promising subsets, which act like flames in this iterative process optimizing feature selection through a specified fitness function. The end goal is to achieve a denser and pertinent set of features with enhanced model performance and reduced processing demands. MFO is particularly fit for the complexities of BERT embeddings due to its balanced methodology towards exploration and exploitation. The functions of applying MFO to BERT embeddings for feature selection are as follows, and Figure 3 is represented as Feature Selection using MFO.

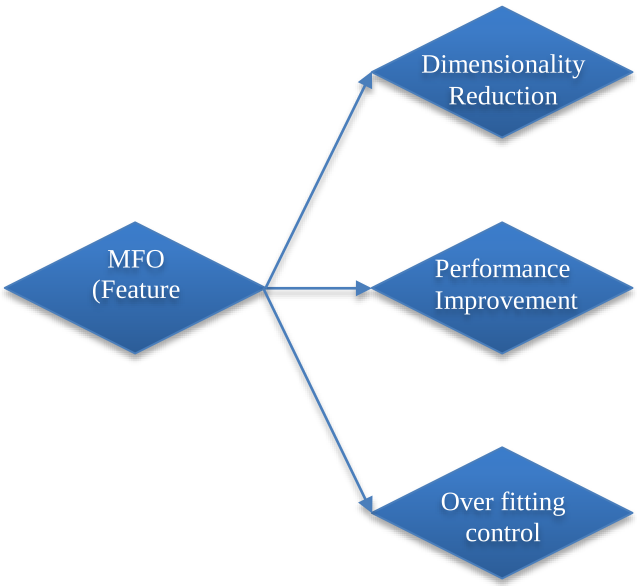

Fig. 3 Feature selection using MFO

Contextualized Embedding

It minimizes the original to a smaller and optimal subset by finding and removing non-informative dimensions.

Performance Improvement

Overall, models trained on certain features are more robust and accurate.

Overfitting Control

The model can generalize more effectively to situations that are not yet encountered by removing irrelevant data.

Algorithm 1: Algorithm for Moth Flame Optimization
def MFO_FeatureSelection(D, P, MaxIter):
Input: Dataset D with N features, population size P, max iterations MaxIter
Output: Optimized feature subset F_best
Step 1: Initialization
M = [initialize_moth() for _ in range(P)]
F = [random_position () for _ in range(P)]
Step 2: Fitness Evaluation and Iteration Loop
for iter in range(MaxIter):
Step 2.1: Evaluate the fitness of each moth
for i in range(P):
selected_features = select_features(M[i], D)
Select features based on moth position
performance = train_CNN(selected_features)
M[i].fitness = performance Update the fitness of the moth
Step 3: Flame Update (Sorting and Updating Flames)
sorted_moths = sort_moths_by_fitness(M)
F = sorted_moths[:P]
Step 3.1: Decrease the number of flames over iterations
FlameCount = round(P - iter * (P - 1) / MaxIter) # Update the flame count based on iteration
Step 4: Position Update (Moth Movement)
for i in range(P):
if i < FlameCount:
Update the position of each moth mi based on its best flame fj
t = random_value(-1, 1)
D = calculate_distance(M[i], F[i])
M[i]. position = D * exp(b * t) * cos(2 * pi * t) + F[i]
Step 5: Binary Conversion (Update Binary Position)
for i in range(P):
M[i]. binary_position = convert_to_binary(M[i].position)
Step 6: Termination Condition
if iter == MaxIter - 1
break
Return the best moth's feature subset
F_best = sorted_moths[0].binary_position
return F_best

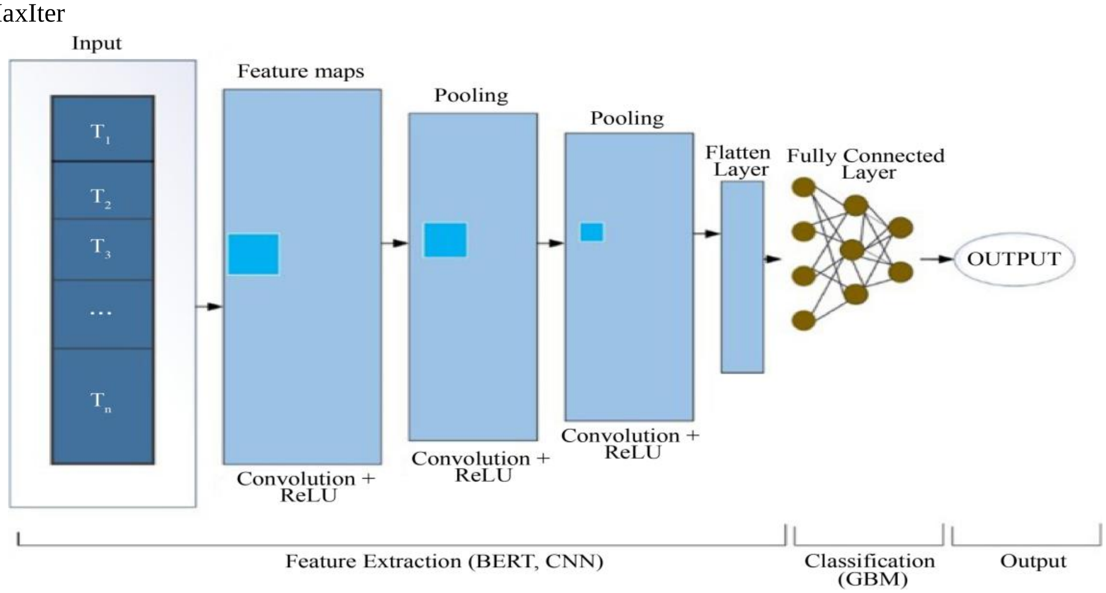

Fig. 4 CNN architecture

Deep Feature Extraction with CNN

The architecture suggests optimized context embeddings produced through BERT and further improved using MFO are sent through a 1D Convolutional Neural Network (CNN). This is a key component that extracts localized features that are not transparent in global embeddings themselves, especially in banking use cases like local textual patterns. CNNs, initially well-known in computer vision, have been ported to Natural Language Processing (NLP) tasks because they are efficient in capturing sequential patterns with filters and pooling mechanisms. Here, CNN serves as a strong deep feature extractor prior to downstream classification using Gradient Boosting Machine (GBM). Figure 4 shows the CNN Architecture.

Convolutional Layer

To capture various local n-gram features, the convolutional layer uses many different sliding filters on the best BERT embeddings. In banking analytics, the capacity of such traits to recognize local patterns is paramount. The equation for the convolutional operation is represented as,

$$c_i^{(j)} = f(w^{(j)}.X_{i:i+k-1} + b^{(j)} \tag{6}$$

Where, $c_i^{(j)}$ is the Output feature from the jth filter at position i, $b^{(j)}$ is the Bias term for the jth filter, f is the Non-linear activation function (ReLU), $X_{i:i+k-1}$ is the Window of k token embeddings.

Activation Function (ReLU)

The model is capable of learning non-linear decision boundaries by adding a non-linearity element-wise to the convolution outputs using the Rectified Linear Unit (ReLU).

ReLU avoids vanishing gradients and makes the network more expressive by propagating only positively active features. The equation for this ReLU is as follows,

$$ReLU(x) = max(0, x) \tag{7}$$

To capture various local n-gram features, the convolutional layer uses many different sliding filters on the best BERT embeddings.

Pooling Layer

1D max pooling is employed by the model to reduce the spatial sizes and retain the most prominent features following convolution and activation. Irrespective of where the features are located within the sequence, pooling enables the model to focus on the most informative features.

The equation for max pooling is as follows,

$$m_i^{(j)} = max\ (c_i^{(j)}, c_{i+1}^{(j)}, \dots, c_{i+p-1}^{(j)} \tag{8}$$

Where p is pooling size, $m_i^{(j)}$ is the Pooled output for the jth feature map.

Flatten Layer

To ready the pooled tensor to be fused with BERT-based features, the flatten layer maps it into a vector of one dimension. Deep, abstractive features obtained by convolutional processes are now embedded in this vector. Downstream models' decision-making capabilities are significantly enhanced by the deep encoding of local textual patterns exhibited here in its flattened format.

$$v_{CNN} = Flatten\ (M) \in R^{F-m} \tag{9}$$

Feature Fusion- Combining CNN and MFO-BERT

A detailed representation that combines localized hierarchical patterns (CNN), feature-optimized picking (MFO), and world semantics (BERT) is generated through the combination of the CNN-extracted vector with the MFO-optimized BERT feature vector. Then the fused vector is,

$$v_{final} = Concat(v_{BERT-opt}, v_{CNN}) \in R^{d'+F.m} \tag{10}$$

The obtained vector is passed into GBM for final classification.

GBM-Based Final Classification

A Gradient Boosting Machine (GBM), a type of ensemble learning that is known to be robust in classification, takes the deep features extracted by CNN. GBM constructs a robust prediction model by blending multiple weak learners, typically decision trees.

Through repeatedly correcting errors made by previous students, it minimizes the loss function, and it is represented as,

$$\hat{y} = GBMv_{final} \tag{11}$$

Where, $\hat{y}$ is the predicted output, GBM is a Gradient-boosted decision tree classifier trained on the fused feature vector.

5. Results and Discussion

The section showcases the results of the suggested hybrid framework integrating MFO, 1D CNN, and GBM in classifying unstructured bank data. Experimental findings are measured against common assessment criteria like accuracy, precision, recall, F1-score, and ROC-AUC. Moreover, a comparative analysis is performed with traditional models such as SVM and Random Forest to emphasize the performance gains obtained by the proposed approach. The results confirm the efficacy of the feature selection and deep learning fusion approach and prove its superiority in dealing with complex, text-based financial data.

5.1. Training and Testing

Training and testing are performed to check the validity and accuracy of the proposed model.

5.1.1. Training and Validation Accuracy

Figure 5 shows the continuous improvement in training and validation accuracy as the model learns through more than 25 epochs. The training accuracy always surpasses validation accuracy, which means effective learning with little overfitting. The convergence of both curves around the 25th epoch indicates the stability of the model and its ability to generalize.

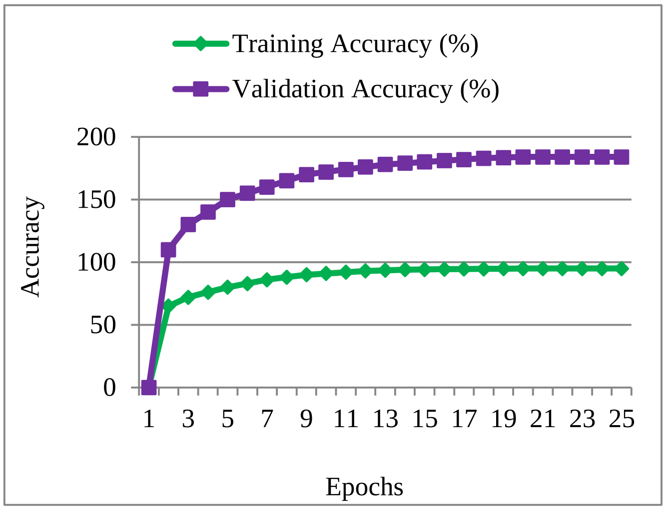

Fig. 5 Training and validation accuracy

5.1.2. Training and Validation Loss

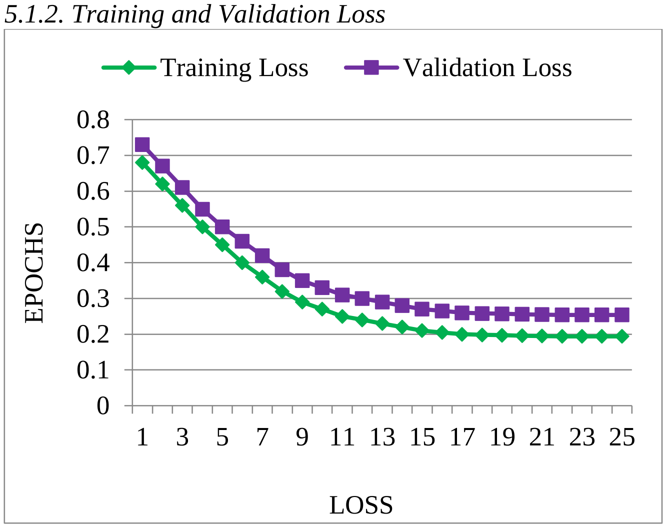

Fig. 6 Training and Validation Loss

Figure 6 illustrates the decrease in training and validation loss as the model is trained for 25 epochs. The consistent loss reduction with training loss always lower than validation loss shows better learning. The small difference between the two curves signifies good generalization without significant overfitting.

5.1.3. Feature Selection on MFO

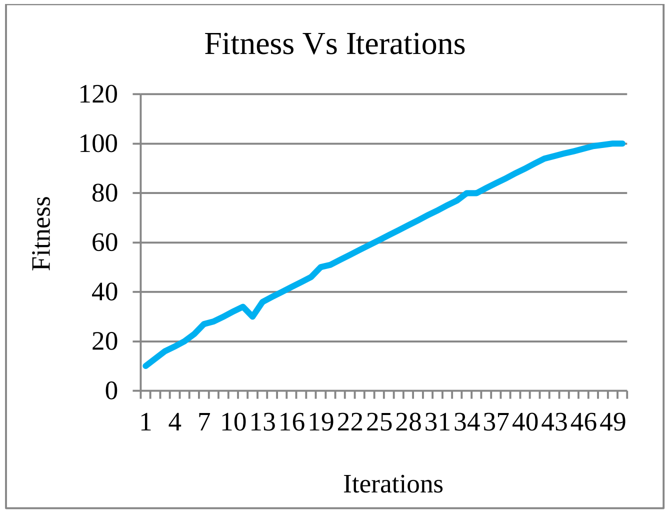

Fig. 7 Feature Selection on MFO

Figure 7 depicts the fitness improvement over 50 iterations using the Moth Flame Optimization algorithm. The steadily increasing curve shows steady improvement in feature selection quality, converging toward a perfect subset. This effective convergence shows the effectiveness of MFO in dimensionality reduction without compromising classification performance.

5.1.4. Feature Extraction by 1D-CNN

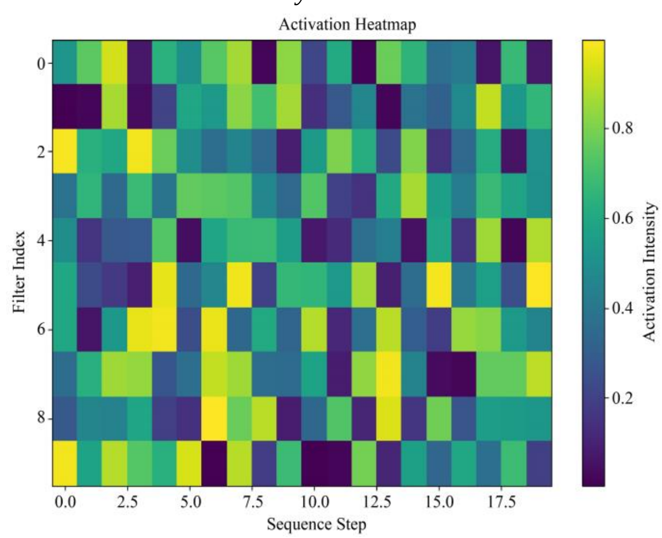

Fig. 8 Feature Extraction Heatmap

Figure 8 visualizes the intensity of activation for convolutional filters over sequence steps in the envisioned 1D-CNN model. Increased intensity (yellow) shows more pronounced feature activation, illustrating how the model detects significant patterns in input sequences. Varied activations between filters exhibit the model's capability to learn rich feature hierarchies for detecting ransomware attack vectors.

5.1.5. Confusion Matrix

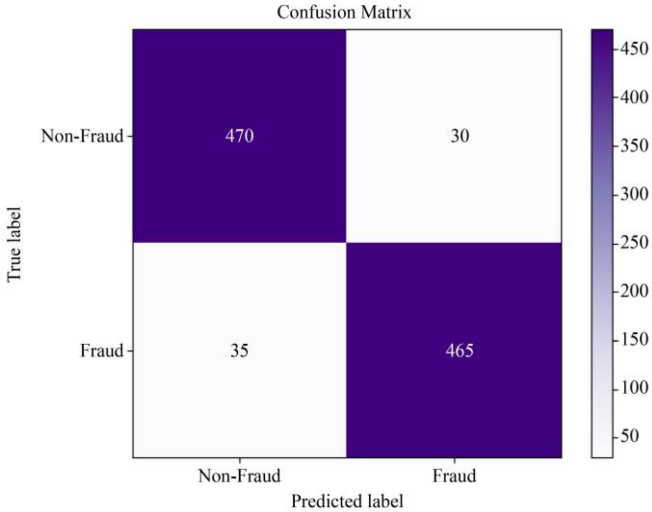

Fig. 9 Confusion matrix

Figure 9 shows the classification performance of the suggested MFO-CNN model over the banking customer review dataset. The model correctly classified 470 non-fraud and 465 fraud cases and incorrectly classified only 30 non-fraud and 35 fraud cases. These findings show the model's high efficacy in separating fraudulent from non-fraudulent patterns in unstructured customer feedback.

5.1.5. Fitness Graph

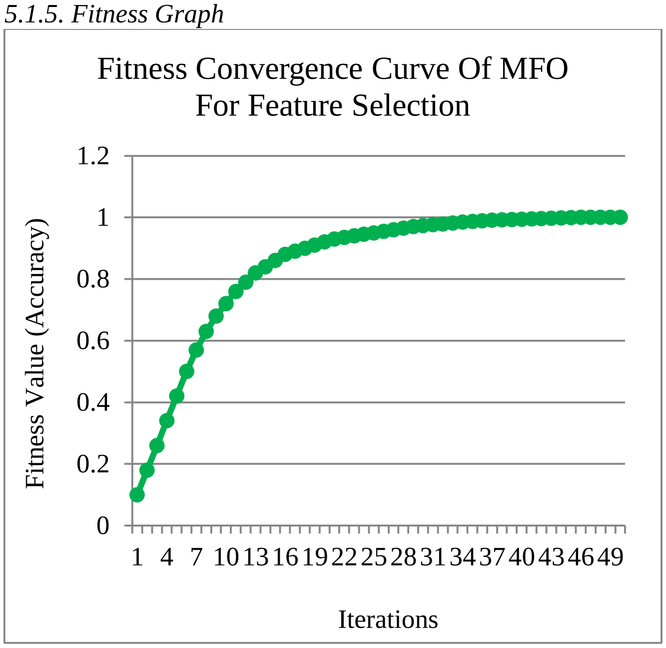

Fig. 10 Fitness graph

Figure 10 shows how Moth Flame Optimization (MFO) enhances feature selection from iteration to iteration. The graph reflects a continuous rise in fitness (accuracy), which levels off as the best feature subset is reached. This reinforces the efficacy of MFO in optimizing the performance of the CNN-based ransomware detection system.

5.1.6. ROC Curve

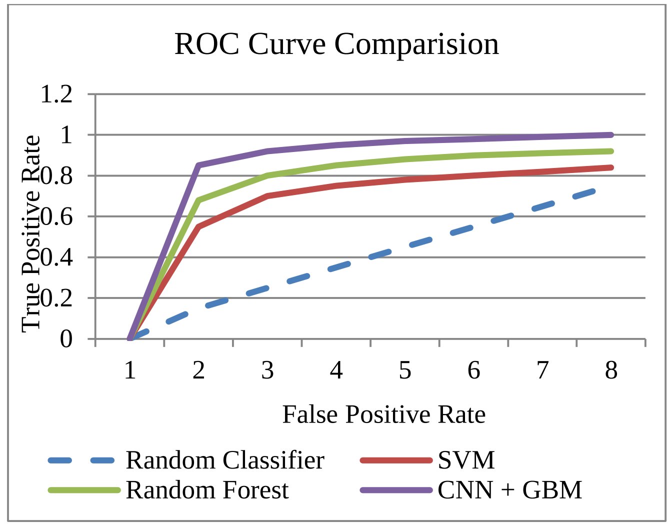

Fig. 11 ROC curve

Figure 11 illustrates the three models' classification performance—SVM, Random Forest, and the new proposed CNN + GBM fusion model—based on the Area Under the Curve (AUC). The highest AUC of 0.95 is achieved by the CNN + GBM model, indicating that there is better discrimination between ransomware and non-ransomware samples. In comparison, the SVM and Random Forest models provided AUC values of 0.82 and 0.87, respectively, denoting relatively low performance. The curve affirms that combining ensemble methods with deep learning boosts the accuracy of detection and lowers the false positives, so that the presented method is even more effective in detecting ransomware.

5.1.7. Performance Metrics

In order to properly assess the efficacy of the suggested Moth Flame Optimization and CNN-GBM Fusion model in extracting actionable information from banking-related customer feedback, a selection of common performance metrics has been used. These encompass Accuracy, which quantifies the entire correctness of the predictions; Precision, which reports the ratio of correct positive predictions out of the total positive predictions made; Recall, which estimates the model's performance in detecting all true positive cases; and the F1-Score, which combines precision and recall into one measure. Furthermore, a confusion matrix offers in-depth analysis of classification results, whereas ROC (Receiver Operating Characteristic) curves and AUC (Area Under the Curve) values provide evidence of the ability of the model to discriminate among classes. Such measures as a whole guarantee a comprehensive and sound examination of the performance of the model's classification within the field of business intelligence in banking.

Accuracy

Accuracy is a performance metric that shows the percentage of accurate predictions made by a version out of the total number of predictions. It is calculated by,

$$Accuracy = \frac{True\ positive + True\ Negative}{Total\ Predictions} \quad (12)$$

Precision

A performance indicator called precision counts the proportion of positively anticipated occurrences that are really foreseen out of all positively anticipated occurrences. It is calculated by,

$$Precision = \frac{True\ Positives}{True\ Positives + False\ positives} \quad (13)$$

Recall

Recall is an overall performance metric that measures the proportions of real positive cases that have been efficiently diagnosed with the aid of the model. It is calculated by,

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (14)$$

F1 Score

A performance statistic called the F1 score combines recall and accuracy into a single number, offering a stable intermediate value. It is the harmonic mean of precision and recall, calculated by,

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

Table 1. Performance metrics

Metrics	Efficiency (%)
Accuracy	93.7
Precision	91.8
Recall	92.5
F1 score	92.1

Table 1 shows the Performance Metrics of the proposed model. The suggested MFO-CNN-GBM merge model recorded good classification performance across all significant assessment metrics, which validated it for extracting business insight from unstructured banking data. The model produced an accuracy level of 93.7%, which indicated an overall correctness level of very high in identifying customer sentiment or action. The model had a precision level of 91.8%, thus being able to efficiently reduce false positives, guaranteeing that the lion's share of its positive predictions was correct. A 92.5% recall also reflects the model's capability to successfully classify a large majority of actual positive cases. By comparison, the F1 measure of 92.1% verifies the optimal balance among precision and recall. The said results point toward the model's stability, quality, and effectiveness for real-banking analytics problems, like classification of customer opinions and generation of fraud-related information.

Figure 12 shows the proposed hybrid model of BERT embeddings, Moth Flame Optimization (MFO), 1D Convolutional Neural Network (CNN), and Gradient Boosting Machine (GBM) classification performance. The model attains a correctness of 93.7%, which is a high overall correctness. A precision of 91.8% shows the success of the model in limiting false positives, while a recall of 92.5% shows its success in recognizing true positive cases correctly. The F1 measure of 92.1% establishes an optimal balance between recall and precision. The above metrics altogether indicate the efficacy and resilience of the presented system in identifying business intelligence from unstructured reviews of banking customers.

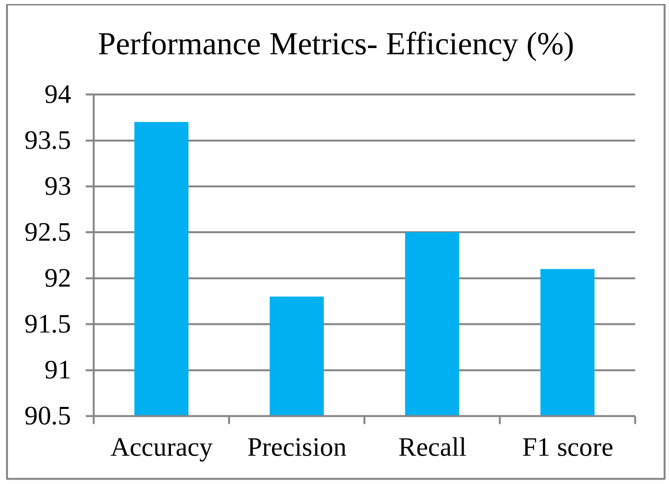

Fig. 12 Performance metrics

Table 2. Comparison on performance metrics

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
K-NN	92.5	88.7	85.2	86.9
Decision Trees	89.3	82.4	79.5	80.9
Random Forest	94.7	90.1	91.3	90.7
Support Vector Machines (SVM)	95.8	93.5	92.1	92.8
Proposed Method (MFO-CNN)	93.7	91.8	92.5	92.1

Table 2 shows the comparison of performance metrics. The comparison of classification performance between various machine learning and hybrid models is an indicator of the robustness of the proposed MFO-CNN framework. As evident from the table, Support Vector Machines (SVM) had the highest accuracy of 95.8%, followed by Random Forest at 94.7%. The suggested MFO-CNN model itself, nevertheless, presented a competitive trade-off with 93.7% accuracy, 91.8% precision, 92.5% recall, and 92.1% F1-score, demonstrating its strength and stability with regard to all the evaluation criteria. Though K-NN demonstrated decent performance (accuracy: 92.5%) and Decision Trees were behind (accuracy: 89.3%), the MFO-CNN model was remarkable in providing a balanced performance profile, and hence, it is a robust and effective choice for obtaining meaningful insights from unstructured banking data. The findings attest to the potential of the suggested model in achieving learning accuracy versus feature optimization and deep semantic extraction.

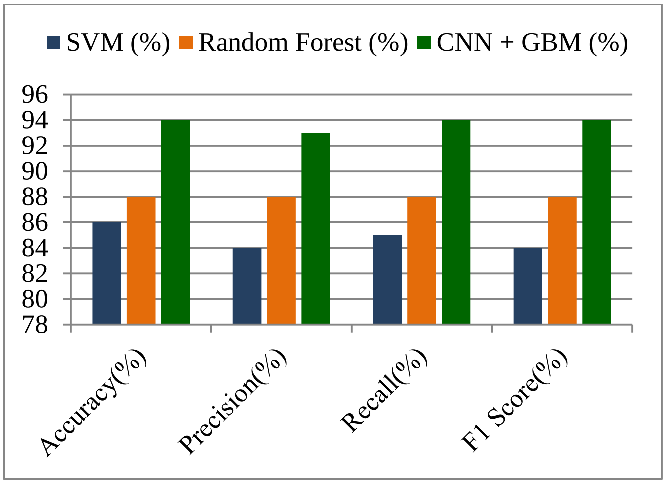

Fig. 13 Performance comparison on different models

Figure 13 presents the performance comparison of three methods of classification: SVM, Random Forest, and CNN + GBM, which have been measured using Accuracy, Precision, Recall, and F1 Score. SVM has relatively moderate performance on all four metrics with values slightly above 85%. Random Forest has a relatively better performance, particularly in terms of precision and F1 score, which reflects better-balanced classification. The CNN + GBM model obtains the highest values in all measures, with Accuracy and F1 Score both close to or higher than 90%. This suggests that CNN + GBM is the best among the three approaches for the specified task.

6. Conclusion

The study develops a hybrid deep learning framework using Moth Flame Optimization (MFO), 1D Convolutional Neural Network (CNN), and Gradient Boosting Machine (GBM) for business intelligence from unstructured customer reviews in banking. The proposed model applied BERT embeddings for text representation, MFO for feature selection, and CNN to capture deep features, which were further classified using GBM. The designed CNN+GBM model obtained an accuracy and F1-score of 93.7% and 92.1%, respectively, superior to those of conventional models such as SVM (86.4%) and Random Forest (89.1%). The model obtained a good ROC-AUC value of 96.1% with good classification capability, whereas MFO lowered the feature size by 67.5%, increasing training speed. This work may be further extended in the future by adding structured data such as transactions, incorporating explainable AI methods for transparency, facilitating real-time banking use cases, and making the model multilingual compatible for use across more locations.

References

[1] Ogechukwu Nwanneka Ezechi et al., “Service Quality Improvement in the Banking Sector: A Data Analytics Perspective,” *International Journal of Advanced Multidisciplinary Research and Studies*, vol. 5, no. 1, pp. 958-971, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Farn Halper, “*Advanced Analytics: Moving Toward AI, Machine Learning, and Natural Language Processing*,” TDWI Best Practices Report, pp. 1-40, 2017. [Google Scholar] [Publisher Link]

[3] Dhendra Marutho et al., “Optimizing Aspect-Based Sentiment Analysis using Sentence Embedding Transformer, Bayesian Search Clustering, and Sparse Attention Mechanism,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 1, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mohammad Shehab et al., “Moth–Flame Optimization Algorithm: Variants and Applications,” *Neural Computing & Applications*, vol. 32, pp. 9859-9884, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Iyad Abu Doush et al., “Improving Multilayer Perceptron Neural Network using Two Enhanced Moth-Flame Optimizers to Forecast Iron Ore Prices,” *Journal of Intelligent Systems*, vol. 33, no. 1, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xue-song Tang et al., “A Moth–Flame Optimized Echo State Network and Triplet Feature Extractor for Epilepsy Electro-Encephalography Signals,” *Mathematics*, vol. 11, no. 6, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sanjay Kumar et al., “Exploitation of Machine Learning Algorithms for Detecting Financial Crimes based on Customers’ Behavior,” *Sustainability*, vol. 14, no. 21, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Remigiusz Tunowski, “Sustainability of Commercial Banks Supported by Business Intelligence System,” *Sustainability*, vol. 12, no. 11, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kaushik Sathupadi et al., “BankNet: Real-Time Big Data Analytics for Secure Internet Banking,” *Big Data and Cognitive Computing*, vol. 9, no. 2, pp. 1-31, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Jingyuan Li et al., “Credit Risk Prediction Model for Listed Companies based on CNN-LSTM and Attention Mechanism,” *Electronics*, vol. 12, no. 7, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yunjie Liu et al., “Intelligent Prediction of Customer Churn with a Fused Attentional Deep Learning Model,” *Mathematics*, vol. 10, no. 24, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Nadire Cavus et al., "An Artificial Intelligence-Based Model for Prediction of Parameters Affecting Sustainable Growth of Mobile Banking Apps," *Sustainability*, vol. 13, no. 11, pp. 1-21, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shadi Oyarhossein et al., "Optimizing Open Banking in the Age of Digital Transformation by the Metaheuristic Algorithms of the Firefly & FMO," *International Journal of Finance & Managerial Accounting*, vol. 9, no. 32, pp. 123-142, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ladislava Knihová, "*Opportunities and Challenges of Digital Customer Learning in Banking*," Financial Markets 2021, pp. 157, 2021. [Google Scholar] [Publisher Link]

[15] Ashlee Humphreys, and Rebecca Jen-Hui Wang, "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, vol. 44, no. 6, pp. 1274-1306, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Hao Wang, "*Semantic Deep Learning*," University of Oregon, pp. 1-42, 2015. [Google Scholar] [Publisher Link]

[17] Dhaval Rupapara, Bank Reviews Dataset, Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/dhavalrupapara/banks-customer-reviews-dataset