Unified BERT-LSTM framework enhances machine learning in fraud detection, financial sentiment, and biomedical classification

Oussama Ndama, Ismail Bensassi, Safae Ndama, El Mokhtar En-Naimi
DSAI2S Research Team, C3S Laboratory, Faculty of Science and Technology Tangier, Abdelmalek Essaâdi University, Tetouan, Morocco

Article Info

Article history:

Received Nov 04, 2024
Revised Oct 22, 2025
Accepted Nov 08, 2025

Keywords:

BERT-LSTM
Biomedical paper abstract classification
Contextual embeddings
Credit card fraud detection
Financial news sentiment analysis
Multi-domain applications
Sequential modeling

ABSTRACT

The current paper proposes a hybrid framework based on the bidirectional encoder representations from transformers (BERT) and long short-term memory (LSTM) networks for classification tasks in three diverse domains: credit card fraud detection (CCFD), financial news sentiment analysis (FNSA), and biomedical paper abstract classification (BPAC). The model leverages the strengths of BERT regarding the learning of contextual embeddings and those of LSTM in capturing sequential dependencies, thus setting the new state-of-the-art performance in each of the three domains. In the CCFD use case, the model was able to achieve an accuracy of 99.11%, considerably outperforming all the competing systems in fraud transaction detection. The BERT-LSTM model achieved a performance of 96.74% for FNSA, improving significantly in sentiment analysis. Finally, the use case of BPAC was robust, with 88.42% accuracy, which clearly classified biomedical abstract sections correctly. It is evident from the findings that this framework generalizes to a wide range of tasks and hence is an adaptable but strong tool in combating challenges of cross-domain classification.

Corresponding Author:

Oussama Ndama
DSAI2S Research Team, C3S Laboratory, Faculty of Science and Technology Tangier
Abdelmalek Essaâdi University
Tetouan, Morocco
Email: oussama.ndama@etu.uae.ac.ma

1. INTRODUCTION

Machine learning (ML) and natural language processing (NLP) have become handy in tackling complex classification tasks across industries. Transformer-based models, such as bidirectional encoder representations from transformers (BERT), have taken a giant leap by enabling machines to grasp rich contextual information from textual data [1], [2]. Unlike older models that process text mainly in one direction, BERT's bidirectional encoding extracts more value from how words relate within a sentence, powering tasks from sentiment analysis to named entity recognition and question answering [3]. Yet many problems also hinge on sequential dependencies and temporal patterns. Long short-term memory (LSTM) networks, with their ability to memorize longer and handle long-distance dependencies in sequential data, are well suited to such settings [4]. LSTMs remain strong options for time series forecasting, speech recognition, and sequence-based text processing, where ordering is essential for accurate prediction [5]. However, tasks requiring both deep contextual understanding and explicit sequence modeling call for models that analyze local context (words or sentences) while verifying ordered patterns. A hybrid approach that incorporates BERT and LSTM is therefore highly useful, exploiting their respective strengths: BERT for contextual understanding and LSTM for sequential pattern capture, making the hybrid particularly effective for complex classification. The proposed work integrates BERT and LSTM into a unified framework and illustrates its application to three diverse domains: credit card fraud detection (CCFD), financial news sentiment analysis (FNSA), and biomedical paper abstract classification (BPAC). Each of these domains has different challenges:

i) CCFD: fraud detection involves studying all transactional patterns over time. Traditional models rely on numeric data, but recent research underlines the importance of textual data, such as a merchant name or a category of transaction that has become a prime source for locating fraud [6]. By combining the best-in-class contextual analysis of BERT and the ability of LSTM to process transaction sequences, there is considerable scope for improving fraud detection rates.
ii) FNSA: in financial markets, news article-based sentiment analysis plays a very crucial role in predicting market trends and investor behaviors [7]. Financial news is filled with nuanced language, often very subtle in nature, where the sentiment may evolve throughout the article. A model that can capture meaning from context and flow, tracking sentiments through the article, provides deep insights far ahead of the sentiment carried by the market.
iii) BPAC: demands the separation of scientific abstracts into pre-defined classes, like background, methods, results, and conclusions. These require relevant understanding in individual sections and how the structure unfolds within the document. The biomedical text is highly dense and technical; hence, closely related sections become tough to distinguish [8]. Such an integrated contextual and sequential information model will be quintessential for the task at hand.

This work aims at showing that a BERT-LSTM hybrid framework generalizes across diverse domains to state-of-the-art performance in each, by factually marrying the strength of understanding meaning and context of words by BERT with the capacity of modeling sequences by LSTM. The proposed framework addresses unique challenges existing in each domain and hence offers a robust solution for cross-domain classification tasks. There are three key contributions of this paper:

i) Unified framework: we hereby propose a hybrid BERT-LSTM model that we can apply in an enormous number of classification tasks; hence, being versatile over various domains.
ii) Cross-domain evaluation: it performs an evaluation on the performance of the framework regarding three different usage scenarios; hence, it can be proved that the network will generalize well and will perform superiorly across different datasets and challenges.
iii) State-of-the-art performance: including both contextual and sequential learning, we present an improved classification performance that provides a new benchmark for each of the studied domains. This work will demonstrate how such hybrid models can solve complex real-world problems by exploiting the strengths of diverse ML architectures. We perform extensive experimentation and evaluation showing how the unified framework provides improved classification performance, along with new avenues for the application of ML and NLP models in diverse fields.

2. RELATED WORKS

Recent advancements in ML have introduced innovative methodologies across CCFD, FNSA, and BPAC, each with unique challenges requiring specialized solutions. In CCFD, techniques target class imbalance and higher detection accuracy. Mienye and Sun [9] proposed a stacking ensemble of LSTM/gated recurrent unit (GRU) enhanced with synthetic minority oversampling technique (SMOTE) and edited nearest neighbor (ENN), achieving high sensitivity and specificity. Benchaji *et al.* [10] integrated attention with LSTM to emphasize critical transactions, improving accuracy. Tang and Liu [11] employed distributed transformer-based knowledge distillation with multi-teacher transfer to boost generalization across financial datasets. Convolutional neural networks (CNNs) have also proven effective for large-scale streams: Mizher and Nassif [12] reported high accuracy, while Ali *et al.* [13] combined SMOTE with CNN/artificial neural network (ANN)/LSTM, with CNNs yielding strong precision and recall.

In FNSA, models are tailored to nuanced financial language that shapes market trends. Mishev *et al.* [14] found BiGRU+attention with word embeddings best for binary headline sentiment. Lim *et al.* [15] similarly highlighted BiGRU+attention for prioritizing salient text segments. Sohangir *et al.* [16] showed CNNs capture sentiment patterns on StockTwits, outperforming several baselines. Atzeni *et al.* [17] adopted a fine-grained approach blending lexical and semantic features, surpassing 72% accuracy. Souma *et al.* [18] linked sentiment to stock price dynamics using LSTMs, and Im *et al.* [19] improved prediction by aggregating headline and content sentiment, underscoring the value of multi-source signals.

BPAC focuses on sentence-level assignment within biomedical abstracts (e.g., background, methods, and results). Gonçalves *et al.* [20] used BiGRU with convolutional layers to capture dependencies, attaining strong precision, recall, and F1 score. Agibetov *et al.* [21] demonstrated FastText can deliver competitive results with low compute on large corpora (PubMed 200k RCT). Banerjee *et al.* [22] leveraged transfer learning pretraining on biomedical text and fine-tuning on smaller targets to show cross-domain portability. Lamurias *et al.* [23] introduced BO-LSTM, using biomedical ontologies to enhance relation classification. Tang *et al.* [24] proposed a ConvLSTM with hierarchical attention for clinical relation extraction, handling class imbalance via structured sequences and attention-based pooling.

These studies highlight the variety of ML approaches tailored for CCFD, FNSA, and BPAC, from ensemble methods and ontology-enhanced LSTM models in CCFD to embedding techniques and attention mechanisms in FNSA, and transfer learning and ontology integration in BPAC. By combining sequential processing, domain knowledge, hierarchical attention, and class balancing, these models have achieved substantial robustness and accuracy in domain-specific text classification, offering valuable methods and insights for handling and interpreting complex data across various fields.

3. RESEARCH METHOD

In this section, we present the development and evaluation of the proposed BERT–LSTM hybrid framework across three use cases. For each domain, textual (and, when applicable, numerical) inputs are processed to yield embeddings that capture rich semantic and temporal relationships. We detail the datasets, preprocessing pipelines, model architecture, training procedures, and evaluation metrics. Using a single framework, we demonstrate that the hybrid approach generalizes across domains and achieves state-of-the-art performance on tasks that require both contextual understanding and sequential modeling.

3.1. Dataset description and preprocessing

The proposed BERT-LSTM hybrid framework was experimented with on three distinct datasets from different domains: CCFD, FNSA, and BPAC. Each dataset needed different preprocessing to feed them to the hybrid model. Details of each dataset and associated preprocessing are shown as follows.

3.1.1. Credit card fraud detection

For the CCFD task, a simulated dataset consisting of 1,296,675 transaction records from 1,000 customers and 800 merchants was used. The dataset included both numerical and textual features, with only 0.60% of the transactions being fraudulent [25]. Among the various features, this dataset includes transaction data, cardholder details, information about merchants, transaction amount, and a binary label on whether a transaction is fraudulent as presented in Table 1.

Table 1. Dataset features

Feature	Description	Type
merchant	The name of the merchant where the transaction occurred	Textual
category	Transaction category (e.g., groceries, fuel)	Textual
amt	Transaction amount	Numeric
is_fraud	Binary label (0=legitimate, 1=fraudulent)	Numeric

The preprocessing steps are as follows:

- Sampling: to reduce computational complexity, a random sample of 100,000 records was selected while maintaining the proportion of fraudulent transactions.
- Handling missing data: for numerical features, missing values were imputed using mean/mode imputation [26]. Most frequent imputation was used for the categorical features.
- Feature engineering: the 'merchant' and 'category' textual features were combined to form one feature that can be embedded by BERT. Numerical features, like 'amt', are scaled for consistency in model training.
- Class balancing: only 0.60% of the transactions were fraudulent; therefore, SMOTE is applied to synthesize new samples for minority class in this problem, thus resolving the imbalanced dataset [27].
- Data splitting: the dataset was split into 80% for training and 20% for validation.

3.1.2. Financial news sentiment analysis

For the FNSA task, a curated dataset of financial news articles was compiled from three major news outlets: CNBC, Reuters, and The guardian. The data spans from 2017 to 2020, covering almost all kinds of economic conditions and market events [28]. After the concatenation of these three sources and excluding any rows with missing values, the final dataset consisted of 53,292 unique news articles. They were categorized as positive, neutral, or negative based on the overall tone and content of the news. Table 2 present a dataset summary.

Table 2. Dataset summary

Source	Time period	No. of articles	Sentiment labels
CNBC	2017-2020	3,080	Positive, neutral, negative
Reuters	2017-2020	32,770	Positive, neutral, negative
The guardian	2017-2020	17,800	Positive, neutral, negative

The preprocessing steps are as follows:

i) Text cleaning:
- The text was converted to lowercase to ensure uniformity.
- Punctuation and non-relevant symbols were removed to reduce noise in the data.
- Common stopwords were filtered out using the natural language toolkit (NLTK) to focus on words that contribute to sentiment classification [29].
- Lemmatization was applied to convert words to their base forms, reducing redundancy and improving model generalization [30].

ii) Polarity analysis using valence aware dictionary and sentiment reasoner (VADER):
- The cleaned text was then analyzed by VADER for polarity. VADER is very suitable for performing sentiment analysis of financial news articles because it measures the emotional valence of the text [31].
- Every news article received a compound score ranging from -1 (very negative) to +1 (very positive), reflecting whether articles were positive, neutral, or negative.

iii) Tokenization:
- The cleaned text was tokenized into sequences of words.
- The BERT tokenizer was used to convert these word sequences into input tokens for the BERT model, ensuring consistent input lengths across all articles [32]. Padding and truncation were applied to adjust input length to a fixed size [33].

iv) Sentiment label encoding:
- The sentiment scores from VADER were used to categorize the sentiment into three labels: positive (compound score>0), neutral (compound score=0), and negative (compound score<0).
- These sentiment categories were then one-hot encoded for multi-class classification, assigning 0 for negative, 1 for neutral, and 2 for positive sentiments.

v) Data splitting: the dataset was divided into two parts: 80% for training and 20% for validation. Each sentiment category was proportionally represented in both the training and validation sets to ensure balanced performance across classes.

3.1.3. Biomedical paper abstract classification

For the BPAC task, we utilized the PubMed 200k RCT dataset, which includes randomized controlled trial abstracts [34]. Each abstract sentence is categorized into one of five predefined sections: background, objective, methods, results, and conclusion. A subset of 500,000 sentences was used for training, ensuring computational efficiency while maintaining the dataset's integrity. Table 3 present a dataset summary.

Table 3. Dataset summary

Source	No. of sentences	Label
Background	45,200	0
Objective	40,300	1
Methods	165,000	2
Results	174,300	3
Conclusion	75,200	4

The preprocessing steps are as follows:

i) Sentence tokenization: sentences were tokenized at both the word and character levels using BERT's tokenizer. The text was padded or truncated to a maximum length of 256 tokens for consistency.

ii) Label encoding: the sentence labels (background, objective, methods, results, conclusion) were one-hot encoded for multi-class classification.

iii) Data splitting: the dataset was split into 80% for training and 20% for validation, ensuring that each sentence category was represented proportionally in both sets.

Each preprocessing pipeline was tailored to domain-specific characteristics. For CCFD, numerical and textual features were processed jointly: missing values were imputed and class imbalance was mitigated using SMOTE. For FNSA, text preprocessing dominated: cleaning, tokenization, and sentiment label encoding. For BPAC, the focus was sentence tokenization and encoding into predefined abstract categories. Table 4 summarizes the key preprocessing steps for all three use cases.

Table 4. Summary of preprocessing steps for all datasets

Step	CCFD	FNSA	BPAC
Sampling	Random sample of 100,000 records	Full dataset of 53,292 articles	Subset of 500,000 sentences
Text cleaning	Combined textual features	Lowercase, punctuation removal, and lemmatization	Tokenization using BERT
Handling missing data	Mean/mode imputation	N/A	N/A
Feature Engineering	Textual features combined, numerical features normalized	N/A	Character-level tokenization
Class balancing	SMOTE applied	N/A	N/A
Label encoding	Binary (fraud/not fraud)	One-hot encoding of sentiment	One-hot encoding of sentence categories
Data splitting	80/20 train-test split	80/20 train-test split	80/20 train-test split

3.2. Model architecture

The BERT–LSTM hybrid framework targets high-complexity classification by integrating BERT's contextual representations with LSTM's sequence modeling. This combination yields a flexible and powerful architecture that generalizes across domains, unifying contextual and sequential learning to produce robust predictions over diverse tasks. Figure 1 illustrates the data flow through the framework and highlights its efficient processing pipeline for accurate inference.

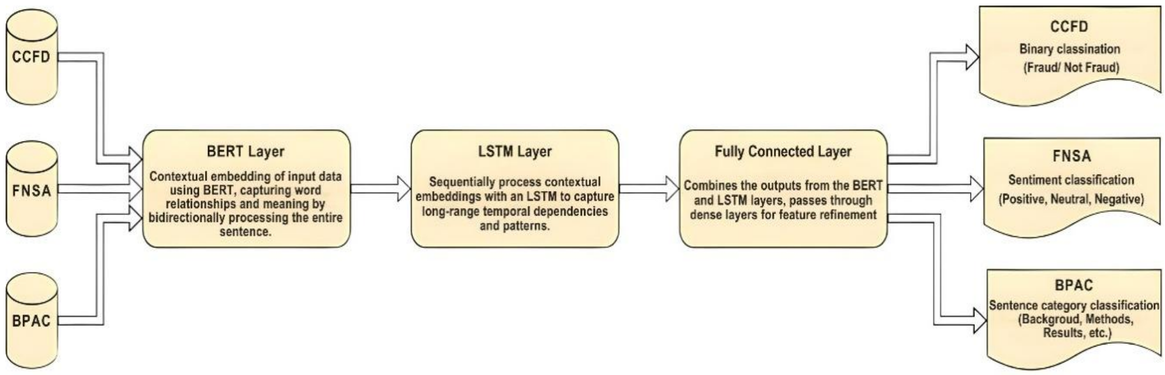

Figure 1. Data flow diagram of the BERT-LSTM framework for multi-domain classification tasks

3.2.1. Bidirectional encoder representations from transformers

BERT is a transformer-based model pre-trained at scale that forms bidirectional representations by jointly conditioning on left and right context across all layers [35]. This enables precise modeling of word meaning and inter-token relations in context, yielding strong performance across diverse NLP tasks [36]. Rooted in the Transformer encoder, BERT relies on self-attention, which allows the model to attend to different parts of the input sequence when making predictions [37].

i) Self-attention mechanism: for each token in the input sequence, BERT computes a weighted sum of the other tokens in the sequence. This is done using the self-attention mechanism, which is mathematically defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Where query (Q), key (K), and value (V) are projections of the input embeddings; d_k is the dimension of the key vector; and the SoftMax function ensures the attention weights sum to 1, allowing the model to assign different weights to different tokens.

ii) Multi-head attention: BERT uses multi-head attention to allow the model to focus on different parts of the sequence simultaneously. It runs multiple attention mechanisms in parallel, each focusing on different aspects of the sequence [38]. The outputs of the attention heads are concatenated and projected to form the final attention output:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (2)$$

where W^O is the projection matrix for the concatenated heads.

iii) Position encoding: since transformers do not have a built-in notion of word order, BERT adds positional encodings to the input embeddings to introduce word position information [39]. The positional encoding is added elementwise to the token embeddings:

$$\text{PE(pos, 2i)} = \sin\left(\frac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (3)$$

$$\text{PE(pos, 2i+1)} = \cos\left(\frac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (4)$$

where pos is the position of the token and i is the dimension.

iv) Feed-forward neural network: after the self-attention layers, BERT uses a fully connected feed-forward neural network for each token [40]. This network is applied independently to each token and consists of two linear transformations with a ReLU activation:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)\, W_2 + b_2 \quad (5)$$

3.2.2. BERT's application across domains

Across the three target domains, BERT serves as a shared contextual text encoder whose inputs and integration with downstream components are adapted to each domain's specific data characteristics and prediction objectives:

i) CCFD: in the CCFD domain, BERT processes textual transaction attributes such as merchant names and category labels, that convey crucial signals about transactional behavior. The pre-trained model encodes these descriptions into contextual embeddings that capture subtle inter-word relations [41]. These embeddings are subsequently fused with numerical features (e.g., amount) for downstream classification, enabling the model to surface linguistic cues indicative of fraud. Figure 2 illustrates BERT's role in the CCFD text-processing branch.

ii) FNSA: in the FNSA domain, BERT is central to processing and contextualizing news articles. Inputs are tokenized with the pre-trained bert-base-uncased tokenizer, with padding/truncation to a fixed length (e.g., 128 tokens) for consistency [42]. The tokenized text is then passed through BERT to produce contextual embeddings that capture nuanced word relations via bidirectional context. These embeddings feed subsequent layers for sentiment classification, leveraging BERT's deep semantic modeling to detect subtle, evolving cues in financial language [43]. Figure 3 illustrates BERT's application to FNSA.

iii) BPAC: in the BPAC domain, BERT models complex word relations in scientific prose. We integrate the pre-trained bert_base_en_uncased, fine-tuned on biomedical text to adapt to domain-specific vocabulary and structure. Abstract sentences are tokenized to a fixed length (256 tokens) with truncation/padding for consistency. BERT then produces contextual embeddings that represent each token relative to its sentence, enabling discrimination among sections such as background, methods, and results. The pooled output passes to a dense layer to refine representations; L2 regularization and dropout are applied to mitigate overfitting and improve generalization [44]. The refined embeddings support sentence-level classification into the predefined abstract categories. Figure 4 depicts BERT's role in biomedical abstract classification.

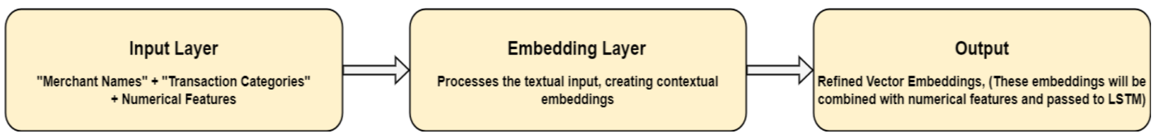

Figure 2. Application of BERT in the processing of textual features from the CCFD domain

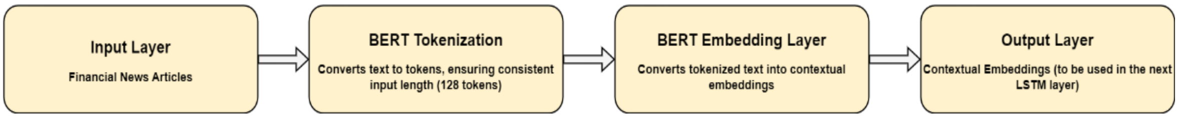

Figure 3. Application of BERT in the processing of textual features from the FNSA domain

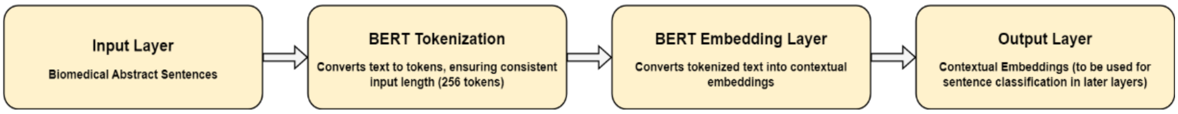

Figure 4. Application of BERT in the processing of textual features from the BPAC domain

3.2.3. Long short-term memory

LSTMs are a type of recurrent neural network (RNN), designed to work on sequential data by maintaining internal memory states over time [45]. This is what allows the LSTM to learn long-range dependencies in the data, making it well suited for tasks demanding temporal sequences [46]. This LSTM contains three main gates, each responsible for controlling how information flows through the network: input, forget, and output gates [47]. These are defined as per the following equations:

i) Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{6}$$

determines how much of the previous state $h_{(t-1)}$ is forgotten.

ii) Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

controls how much of the new information x_t is written into the internal state.

iii) Cell state update:

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{8}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{9}$$

updates the cell state C_t based on the forget gate and input gate operations.

iv) Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

Determines the output based on the current cell state:

$$h_t = o_t \times \tanh(C_t) \tag{11}$$

This architecture enables the LSTM to capture information given extremely long sequences, which makes them very suitable for financial transaction data, time-dependent sentiment in news articles, and sentence classification in scientific abstracts.

3.2.4. LSTM application across domains

Across the three target domains, LSTM layers operate as the shared sequential backbone, ingesting BERT-derived embeddings and modeling temporal dependencies tailored to each task's dynamics:

i) CCFD: in the CCFD domain, BERT-derived embeddings feed an LSTM tailored for sequential data. Over transaction sequences, the LSTM learns temporal dependencies e.g., periodic merchant categories or timing patterns by maintaining internal states that retain past information critical for detecting evolving fraud. Order and timing cues thus become signals for identifying anomalous patterns [48]. The LSTM output is passed to a fully connected layer for binary classification (fraudulent vs. legitimate), with a sigmoid activation mapping to fraud likelihood. With BERT handling contextual transaction details and LSTM modeling sequence dynamics, the hybrid yields stronger fraud detection. Figure 5 illustrates the LSTM stage following BERT in the CCFD pipeline.

ii) FNSA: in FNSA, the LSTM complements BERT by modeling temporal dependencies that contextual embeddings only partially capture. BERT's article embeddings are fed to the LSTM, which tracks sentiment flow across the narrative e.g., shifts from neutral to positive or positive to negative by retaining information from earlier segments via its internal states [49]. The LSTM output is passed to a fully connected layer, with a SoftMax output over three classes (positive, neutral, and negative). This BERT–LSTM architecture exploits deep context and sequence dynamics, yielding more accurate sentiment classification for financial news. Figure 6 depicts the integration of LSTM with BERT for FNSA.

iii) BPAC: in BPAC, LSTM augments BERT's contextual embeddings by modeling sequential dependencies within and across abstract sentences. BERT produces token-level representations that serve as inputs to the LSTM, which captures temporal relations among words and preserves the narrative order of sentences, crucial in scientific prose where ordering conveys meaning [50]. The LSTM output is combined with auxiliary features (e.g., sentence position) and passed to a dense layer with SoftMax to classify sentences into background, objective, methods, results, or conclusion. This BERT–LSTM integration better captures abstract structure for accurate section labeling. Figure 7 illustrates the LSTM stage following BERT in the BPAC model.

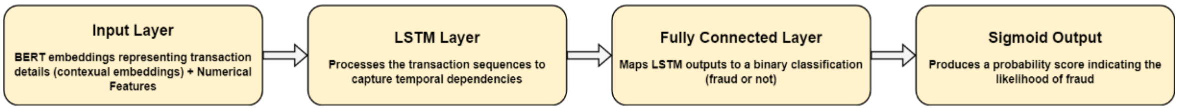

Figure 5. Application of LSTM for sequential processing in CCFD model

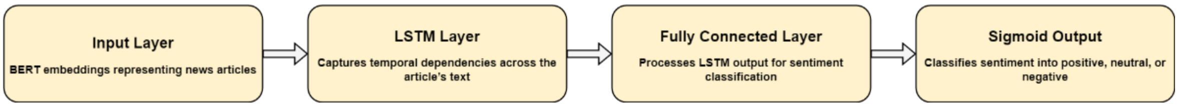

Figure 6. Application of LSTM for sequential processing in FNSA model

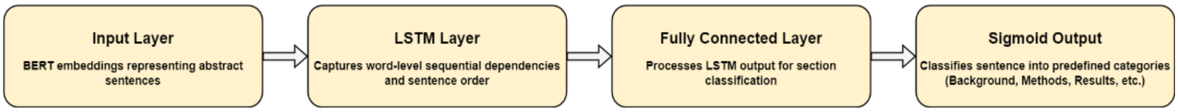

Figure 7. Application of LSTM for sequential processing in BPAC model

3.2.2. Integration and output of BERT-LSTM hybrid model across domains

In the final stage of our model architecture, we integrate outputs from BERT and LSTM layers to generate tailored, domain-specific predictions. This integration capitalizes on BERT's strength in contextual understanding and LSTM's ability to model sequential dependencies, creating a cohesive representation that boosts accuracy and reliability across tasks. This approach also makes the model adaptable to the unique demands of each domain.

The integration process begins by merging the contextual embeddings generated by BERT with the sequential patterns captured by LSTM. This combination enables the model to understand both word meanings within context and the sequential relationships in the data crucial for tasks like fraud detection, sentiment analysis, and scientific text classification. Once the combined embeddings are formed, they pass through one or more fully connected (dense) layers. These layers refine the features extracted by BERT and LSTM, transforming them into compact, domain-specific representations. This dense layer aids the model in learning complex patterns and relationships within the data, enhancing classification accuracy.

Each domain uses a unique output layer configuration and activation function to meet its specific classification requirements:

– CCFD: a sigmoid activation function in the output layer produces a probability score between 0 and 1, indicating the likelihood of fraud. This score enables binary classification, flagging transactions as either fraudulent or legitimate based on risk.

– FNSA: a SoftMax activation function provides a probability distribution over three sentiment categories positive, neutral, and negative. The model classifies news articles based on the highest probability, enabling nuanced sentiment analysis for financial news.
– BPAC: here, a SoftMax activation function generates probabilities across predefined abstract categories (methods, results, and conclusion). The model then classifies each sentence based on the highest probability, helping structure the abstract's content.

To achieve optimal learning and generalization, the model employs task-specific loss functions and optimization techniques:

– Loss functions: for binary classification in CCFD, binary cross-entropy is used, providing precise measurements of model accuracy in fraud detection. For multi-class tasks in FNSA and BPAC, categorical cross-entropy is applied, evaluating the model's effectiveness in sentiment and sentence classification [51].
– Optimization strategies: techniques like learning rate scheduling, dropout, and early stopping are integrated to minimize overfitting and enhance model performance [52]. Learning rate scheduling adjusts learning rates for efficient convergence, dropout mitigates overfitting by temporarily disabling neurons during training, and early stopping halts training when the validation loss ceases to improve, avoiding excessive model complexity.

In summary, the integration and output phase merges BERT and LSTM outputs for domain-specific classifications, creating a robust and adaptable architecture. By unifying contextual and sequential learning, the model effectively serves various NLP tasks, achieving accurate, domain-tailored predictions across different applications.

3.3. Evaluation approach

This section details the methods employed to evaluate the BERT-LSTM model's performance across multiple domains. Rigorous evaluation is essential to confirm the model's capability in tasks like fraud detection, sentiment analysis, and scientific text classification. By choosing appropriate evaluation metrics, designing experimental setups, and leveraging high-performance hardware, we ensure that the model's results are both reliable and generalizable [53]. The performance of the BERT-LSTM model is evaluated using specific metrics tailored to each domain:

– Accuracy measures the proportion of correct predictions among the total number of predictions. It is defined as:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \tag{12}$$

– Precision measures the proportion of true positive predictions among the total predicted positives. This metric is especially relevant in fraud detection to reduce false alarms. It is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}} \tag{13}$$

– Recall or sensitivity, measures the proportion of true positive predictions among all actual positive instances. In fraud detection, high recall helps minimize missed fraudulent transactions. The formula is:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}} \tag{14}$$

– The F1 score is the harmonic mean of precision and recall, providing a balanced metric for imbalanced datasets. It is especially useful in fraud detection and sentiment analysis, where precision and recall may need to be balanced. The formula is:

$$\text{F1} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{15}$$

Each metric serves a specific purpose across domains: for CCFD, metrics such as accuracy, precision, recall, and F1 score are used to measure the model's effectiveness in distinguishing fraudulent from legitimate transactions, with a focus on minimizing false negatives. In FNSA, accuracy, precision, recall, and F1 score are employed to evaluate sentiment classification accuracy across positive, neutral, and negative sentiments. For BPAC, multi-class classification metrics, including accuracy, recall, and F1 score, are essential for assessing the model's ability to correctly categorize sentences into abstract sections.

4. RESULTS AND DISCUSSION

This section presents the performance results of the BERT-LSTM model over three domains: CCFD, FNSA, and BPAC. We discuss these results in the context of the previously outlined evaluation metrics and provide interpretations about the strengths and limitations of the model. On the domain of CCFD, fraud detection by the BERT-LSTM model outperformed other models with an accuracy of 99.11%, a precision of 98.27%, a recall of 100%, and an F1 score of 99.13%. These results are shown in Table 5, which reflects that the model was highly accurate, in fact, perfect in recall for fraudulent transaction detection, which means all fraud cases were correctly identified.

Table 5. Performance metrics of the BERT-LSTM hybrid model in CCFD domain

Metric	Value
Accuracy	0.9911
Precision	0.9827
Recall	1.0000
F1 score	0.9913

Figure 8 gives a breakdown of the performance for each class in detail. Based on this value, for the class 'Non-fraud', the model gave a result as precision-100%, F1 score-0.99, and correctly predicted as 19,758. Whereas the fraud class had a precision and recall of 98% and 100%, respectively, with 20,003 instances in the support set, hence proved reliable on various occasions with minimum false negatives. This is important because, essentially in interpretation, the model's high recall for fraud cases means it missed no fraudulent transactions. The high F1 score and precision of both classes speak to the model's ability to correctly classify transactions with extremely minimal error. Contextual embeddings from BERT allow for a better representation of nuanced patterns in transaction details compared to traditional models. Its LSTM component then captures temporal dependencies of sequences in these transactions. However, there is always a trade-off between precision and recall-the very high recall must trigger off a small number of false positives.

	precision	recall	f1-score	support
Non-Fraud	1.00	0.98	0.99	19758
Fraud	0.98	1.00	0.99	20003
accuracy			0.99	39761
macro avg	0.99	0.99	0.99	39761
weighted avg	0.99	0.99	0.99	39761

Validation Loss: 0.0375

Figure 8. Classification report of the BERT-LSTM hybrid model in CCFD domain

The BERT-LSTM model for sentiment classification in FNSA achieved an overall accuracy of 96.74%, precision of 97.09%, recall of 96.64%, and an F1 score of 96.67%, as depicted in Table 6. This model obtains strong results across the sentiment categories such as negative, neutral, and positive, as depicted by the per-class metrics in Figure 9. Its performance is exceptionally fine in determining neutral sentiments maybe because there will be clearer linguistic markers in the financial news dataset. In summary, high precision and F1 scores within the sentiment classes indicate that BERT-LSTM performs quite well in capturing subtle sentiment cues in financial text. The high performance of the Neutral class strengthens the belief that BERT embeddings handle the complex vocabulary and expressions in financial news effectively, while LSTM tackles the sequence of sentiment flow within the text. The relatively lower performance of the model in Positive sentiment detection can thus be attributed to the possibility that certain features of its language might have overlapped with those in the neutral class and, therefore, may require more fine-tuning of the BERT embeddings on domain-specific financial corpora. In the domain of BPAC, the model achieved 88.42% accuracy, 88.53% precision, 88.42% recall, and an F1 score of 88.45%. Figure 10 presents detailed metrics for each abstract section; indeed, the performance is varied.

Table 6. Performance metrics of the BERT-LSTM hybrid model in FNSA

Metric	Value
Accuracy	0.9674
Precision	0.9709
Recall	0.9664
F1 score	0.9667

	precision	recall	f1-score	support
Negative	0.97	0.96	0.96	1870
Neutral	0.99	0.97	0.98	1827
Positive	0.95	0.97	0.96	1633
accuracy			0.97	5330
macro avg	0.97	0.97	0.97	5330
weighted avg	0.97	0.97	0.97	5330

Figure 9. Classification report of the BERT-LSTM Hybrid Model in FNSA domain

	precision	recall	f1-score	support
BACKGROUND	0.722143	0.759294	0.740253	2663.000000
CONCLUSIONS	0.866137	0.910755	0.887885	4426.000000
METHODS	0.936133	0.934981	0.935557	9751.000000
OBJECTIVE	0.713362	0.696256	0.704705	2377.000000
RESULTS	0.927255	0.900545	0.913705	10276.000000
accuracy	0.884244	0.884244	0.884244	0.884244
macro avg	0.833006	0.840366	0.836421	29493.000000
weighted avg	0.885259	0.884244	0.884549	29493.000000

Figure 10. Classification report of the BERT-LSTM hybrid model in BPAC domain

As an interpretation, very high performance of the model in methods and results sections manifests its capabilities in capturing the distinct scientific terminologies within these categories. However, the poor scores in background and objective indicate that there were difficulties in these classes due to the overlapping between them. These contextual embeddings from BERT, underlined with sequential processing by LSTM, enabled the identification of the sections effectively, but further improvements could have been achieved using features such as the sentence position of the abstract.

Compared with traditional methods in multi-domain applications, all the results show great improvements in both classification accuracy and robustness. On CCFD, the recall of the single model of BERT-LSTM outperforms that of others in general accuracy, while the stacking ensemble including LSTM and GRU by Mienye and Sun [9] achieved only a 90.5% recall. For example, the approach by Mishev *et al.* [14] combining LSTM and UMAP achieved an accuracy of 96.7% with a recall of 91.9%, compared to 93.5% accuracy achieved with the use of CNN by Mizher and Nassif [12]. By combining BERT with its contextual embeddings, the BERT-LSTM model gives more enhanced representations of transaction features, which thus now enhances fraud detection. For instance, in FNSA, the BERT-LSTM model outperformed previous approaches, including the CNN model proposed by Sohangir *et al.* [16], with an accuracy of 90.93% and F1 score of 90.86%. Similarly, while models such as the BiGRU+attention with GloVe embeddings proposed by Mishev *et al.* [14] and the bidirectional LSTM encoder-decoder by Lim *et al.* [19] were competitive, they lacked deep context sensitivity, which BERT-LSTM is introducing into sentiment analysis. This is most evident in neutral sentiment classification, where traditional models often struggle with complex financial terminology. By pairing BERT's deep contextual representations with LSTM's sequential processing, the BERT–LSTM hybrid more accurately captures sentiment shifts across financial topics. For BPAC, the model attains an average F1 score of 88.45%, outperforming traditional LSTM-based systems that fail to separate overlapping sections in scientific abstracts. While some methods excel within a single domain, elevating BERT's context-aware embeddings through LSTM's sequence modeling improves adaptability and accuracy across diverse textual structures.

This comparison highlights the contribution of the BERT–LSTM architecture. BERT embeddings provide rich, context-sensitive representations that raise performance across domains, while LSTM enables learning from temporal/structured inputs essential to fraud detection, sentiment analysis, and scientific text segmentation. The model's cross-domain versatility reflects the synergy between BERT's contextual depth and LSTM's sequence handling. Limitations remain: in CCFD, balancing precision and recall is difficult when fraudulent transactions closely resemble legitimate ones; in FNSA, positive sentiment detection shows slight degradation, suggesting domain-specific tuning; in BPAC, overlap between background and objective hampers separation. Finally, the approach is computationally demanding, raising scalability concerns for real-time deployment in resource-constrained settings.

5. CONCLUSION

This work presents a unified BERT-LSTM hybrid framework that has obtained state-of-the-art performance in complex classification tasks with multi-domain applications. Experiments are conducted on the CCFD, FNSA, and BPAC datasets to highlight this. Our proposed framework unifies these by incorporating deep contextual embeddings from BERT with the sequential processing in LSTM to better capture complex relationships among data. This enhances the performance on state-of-the-art methods that capture subtle categories, such as the distinction between legitimate versus fraudulent transactions, sentiment in financial language, and the structural segments within scientific abstracts. The flexibility and power of this architecture can be seen in these results. In CCFD, the model showed better accuracy and recall, underpinning its capability of detecting subtle fraud patterns without compromising balanced performance in classifying valid transactions. The unified framework FNSA did exceptionally well in recognizing complex financial sentiments, especially neutral sentiment, where previous models many times go off the mark. The model yielded an outstanding F1 score in BPAC by accurately separating overlapping abstract sections, evidence of its capability to process both contextual and sequential information. This all-in-one framework sets a strong basis for tasks needing the comprehension of linguistic input, added with temporally specified information, and can be adapted easily to various domains in NLP. While performing, there remain a few challenges. In CCFD, the challenge is to get a good balance between precision and recall in detecting fraudulent patterns, which resemble similar characteristics to ordinary transactions. For FNSA, model performance results on positive sentiment detection indicate domain-specific tuning is needed. In BPAC, background/objective cannot easily be differentiated due to the content overlap. Besides, this might pose scalability challenges when considering real-time applications, particularly in resource-constrained environments, due to the high computational complexity of the model BERT-LSTM. In summary, this suggests that the mutual benefit of contextual depth from BERT and sequence processing strength from LSTM in an integrated framework results in an effective performance of the BERT-LSTM model across many domains. It thus brings invaluable insights into the development of classification models across a wide array of fields, showing how hybrid models can improve both accuracy and robustness and be more adaptive in complex NLP tasks across multi-domain applications.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Oussama Ndama	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Ismail Bensassi						✓				✓				
Safae Ndama						✓				✓				
El Mokhtar En-Naimi		✓		✓		✓				✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

All datasets used in this study are publicly accessible and sourced from well-established repositories. Details for each experiment are provided as follows:

– The data that support the findings for the credit card fraud detection (CCFD) experiments are openly available on GitHub at https://github.com/namebrandon/Sparkov_Data_Generation

– The data that support the findings for the financial news sentiment analysis (FNSA) experiments are openly available on Kaggle at https://www.kaggle.com/datasets/notlucasp/financial-news-headlines
– The data that support the findings for the biomedical paper abstract classification (BPAC) experiments are openly available in the ACL Anthology at https://www.aclweb.org/anthology/I17-2052.pdf, reference [34].

REFERENCES

[1] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[2] J. Parmar, S. Chouhan, V. Raychoudhury, and S. Rathore, "Open-world machine learning: applications, challenges, and opportunities," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–37, Sep. 2022, doi: 10.1145/3561381.

[3] J. Singh, *Natural language processing in the real world: text processing, analytics, and classification*. New York: Chapman and Hall/CRC, 2023, doi: 10.1201/9781003264774.

[4] F. Cheng and Y. Miyao, "Classifying temporal relations by bidirectional LSTM over dependency paths," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Jul. 2017,pp. 1-6, doi: 10.18653/v1/p17-2001.

[5] R. Huang *et al.,* "Well performance prediction based on long short-term memory (LSTM) neural network," *Journal of Petroleum Science and Engineering*, vol. 208, Oct. 2021, doi: 10.1016/j.petrol.2021.109686.

[6] P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, "Fraud detection with natural language processing," *Machine Learning*, vol. 113, pp. 5087–510, Jul. 2023, doi: 10.1007/s10994-023-06354-5.

[7] M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, "Machine learning sentiment analysis, COVID-19 news and stock market reactions," *Research in International Business and Finance*, vol. 64, Jan. 2023, doi: 10.1016/j.ribaf.2023.101881.

[8] S. Zhao, C. Su, Z. Lu, and F. Wang, "Recent advances in biomedical literature mining," *Briefings in Bioinformatics*, vol. 22, no. 3, Mar. 2020, doi: 10.1093/bib/bbaa057.

[9] I. D. Mienye and Y. Sun, "A deep learning ensemble with data resampling for credit card fraud detection," *IEEE Access*, vol. 11, pp. 30628–30638, Mar. 2023, doi: 10.1109/access.2023.3262020.

[10] I. Benchaji, S. Douzi, B. E. Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00541-8.

[11] Y. Tang and Z. Liu, "A distributed knowledge distillation framework for financial fraud detection based on transformer," *IEEE Access*, vol. 12, pp. 62899–62911, Apr. 2024, doi: 10.1109/access.2024.3387841.

[12] M. Z. Mizher and A. B. Nassif, "Deep CNN approach for unbalanced credit card fraud detection data," in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, Feb. 2023, pp. 1-7, doi: 10.1109/aset56582.2023.10180615.

[13] M. N. Y. Ali, T. Kabir, N. L. Raka, S. S. Toma, M. L. Rahman, and J. Ferdaus, "SMOTE based credit card fraud detection using convolutional neural network," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*, Dec. 2022, pp. 55-60, doi: 10.1109/iccit57492.2022.10054727.

[14] K. Mishev *et al.*, "Performance evaluation of word and sentence embeddings for finance headlines sentiment analysis," in *Communications in Computer and Information Science*, 2019, pp. 161–172, doi: 10.1007/978-3-030-33110-8_14.

[15] S. L. O. Lim, H. M. Lim, E. K. Tan, and T.-P. Tan, "Examining machine learning techniques in business news headline sentiment analysis," in *Computational Science and Technology*, 2019, pp. 363–372, doi: 10.1007/978-981-15-0058-9_35.

[16] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big data: Deep learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, Jan. 2018, doi: 10.1186/s40537-017-0111-6.

[17] M. Atzeni, A. Dridi, and D. R. Recupero, "Fine-grained sentiment analysis on financial microblogs and news headlines," in *Communications in Computer and Information Science*, 2017, pp. 124–128, doi: 10.1007/978-3-319-69146-6_11.

[18] W. Souma, I. Vodenska, and H. Aoyama, "Enhanced news sentiment analysis using deep learning methods," *Journal of Computational Social Science*, vol. 2, no. 1, pp. 33–46, Feb. 2019, doi: 10.1007/s42001-019-00035-x.

[19] T. L. Im, P. W. San, C. K. On, R. Alfred, and P. Anthony, "Impact of financial news headline and content to market sentiment," *International Journal of Machine Learning and Computing*, vol. 4, no. 3, pp. 237–242, Mar. 2014, doi: 10.7763/ijmlc.2014.v4.418.

[20] S. Gonçalves, P. Cortez, and S. Moro, "A deep learning classifier for sentence classification in biomedical and computer science abstracts," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6793–6807, Jul. 2019, doi: 10.1007/s00521-019-04334-2.

[21] A. Agibetov, K. Blagec, H. Xu, and M. Samwald, "Fast and scalable neural embedding models for biomedical sentence classification," *BMC Bioinformatics*, vol. 19, no. 1, Dec. 2018, doi: 10.1186/s12859-018-2496-4.

[22] S. Banerjee, D. K. Sanyal, S. Chattopadhyay, P. K. Bhowmick, and P. P. Das, "Segmenting scientific abstracts into discourse categories: a deep learning-based approach for sparse labeled data," in *JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, Aug. 2020, pp. 429–432, doi: 10.1145/3383583.3398598.

[23] A. Lamurias, D. Sousa, L. A. Clarke, and F. M. Couto, "BO-LSTM: classifying relations via long short-term memory networks along biomedical ontologies," *BMC Bioinformatics*, vol. 20, no. 1, Jan. 2019, doi: 10.1186/s12859-018-2584-5.

[24] L. Tang, F. Teng, Z. Ma, L. Huang, M. Xiao, and X. Li, "Convolutional LSTM network with hierarchical attention for relation classification in clinical texts," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, vol. 15, pp. 1–8, doi: 10.1109/ijcnn.2019.8852281.

[25] E. Leroy, "Synthetic credit card transaction generator used in the Sparkov program," *GitHub*, 2024. Accessed: Jul 31, 2024. [Online]. Available: https://github.com/namebrandon/Sparkov_Data_Generation

[26] A. Desiani, N. R. Dewi, A. N. Fauza, N. Rachmatullah, M. Arhami, and M. Nawawi, "Handling missing data using combination of deletion technique, mean, mode and artificial neural network imputation for heart disease dataset," *Science & Technology Indonesia*, vol. 6, no. 4, Oct. 2021, doi: 10.26554/sti.2021.6.4.303-312.

[27] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[28] L. Pham, "Financial news headlines data." *Kaggle.* 2020. Accessed: Jul 31, 2024. [Online]. Available: https://www.kaggle.com/datasets/notlucasp/financial-news-headlines

[29] M. Wang and F. Hu, "The application of NLTK library for Python natural language processing in corpus research," *Theory and Practice in Language Studies*, vol. 11, no. 9, pp. 1041–1049, Sep. 2021, doi: 10.17507/tpls.1109.09.

[30] R. Pramana, N. Debora, J. J. Subroto, A. A. S. Gunawan, and N. Anderies, "Systematic literature review of stemming and lemmatization performance for sentence similarity," in *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, Nov. 2022, doi: 10.1109/icitda55840.2022.9971451.

[31] C. J. Hutto and E. Gilbert, "VADER: a parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, May 2014, vol. 8, no. 1, pp. 216–225, doi: 10.1609/icwsm.v8i1.14550.

[32] A. Bello, S.-C. Ng, and M.-F. Leung, "A BERT framework to sentiment analysis of tweets," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010506.

[33] J. Kamps, N. Kondylidis, and D. Rau, "Impact of tokenization, pretraining task, and transformer depth on text ranking," in *29th Text REtrieval Conference, TREC 2020*, Jan. 2021, pp. 1-8.

[34] F. Dernoncourt and J. Y. Lee, "PubMed 200k RCT: a dataset for sequential sentence classification in medical abstracts," in *International Joint Conference on Natural Language Processing*, Oct. 2017, vol. 2, pp. 308–313.

[35] F. Miletic and S. S. I. Walde, "A systematic search for compound semantics in pretrained BERT architectures," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, Jan. 2023, pp. 1499–1512, doi: 10.18653/v1/2023.eacl-main.110.

[36] E. C. G. -Merchan, R. G. -Brizuela, and S. G. -Carvajal, "Comparing BERT against traditional machine learning models in text classification," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 4, Apr. 2023, pp. 352-356, doi: 10.47852/bonviewjcce3202838.

[37] Y. Hao, L. Dong, F. Wei, and K. Xu, "Self-attention attribution: interpreting information interactions inside Transformer," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 14, pp. 12963–12971, May 2021, doi: 10.1609/aaai.v35i14.17533.

[38] X. Zhang, Z. Wu, K. Liu, Z. Zhao, J. Wang, and C. Wu, "Text sentiment classification based on BERT embedding and sliced multi-head self-attention Bi-GRU," *Sensors*, vol. 23, no. 3, Jan. 2023, doi: 10.3390/s23031481.

[39] A. Kazemnejad, I. Padhi, N. R. Karthikeyan, P. Das, and S. Reddy, "The impact of positional encoding on length generalization in Transformers," in *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS 2023)*, Dec. 2023, pp. 24892-24928.

[40] J. Bona-Pellissier, F. Bachoc, and F. Malgouyres, "Parameter identifiability of a deep feedforward ReLU neural network," *Machine Learning*, vol. 112, no. 11, pp. 4431–4493, Aug. 2023, doi: 10.1007/s10994-023-06355-4.

[41] Z. Gou and Y. Li, "Integrating BERT embeddings and BiLSTM for emotion analysis of dialogue," *Computational Intelligence and Neuroscience*, vol. 2023, pp. 1–8, May 2023, doi: 10.1155/2023/6618452.

[42] M. P. Geetha and D. K. Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model," *International Journal of Intelligent Networks*, vol. 2, pp. 64–69, Jan. 2021, doi: 10.1016/j.ijin.2021.06.005.

[43] S. Mittal, A. Chauhan, and C. K. Nagpal, "Stock market prediction by incorporating news sentiments using Bert," in *Studies in Computational Intelligence*, 2022, pp. 35–45, doi: 10.1007/978-3-030-96634-8_4.

[44] M. Li, "A practical significant technic in solving overfitting: Regularization," *Theoretical and Natural Science*, vol. 5, no. 1, pp. 253–258, May 2023, doi: 10.54254/2753-8818/5/20230433.

[45] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D Nonlinear Phenomena*, vol. 404, Mar. 2020, doi: 10.1016/j.physd.2019.132306.

[46] R. DiPietro and G. D. Hager, "Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, 2020, pp. 503–519, doi: 10.1016/b978-0-12-816176-0.00026-0.

[47] A. Pulver and S. Lyu, "LSTM with working memory," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 845-851, doi: 10.1109/ijcnn.2017.7965940.

[48] S. Liu, G. Liao, and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, May 2018, pp. 2787-2790, doi: 10.1109/iciea.2018.8398183.

[49] S. Usmani and J. A. Shamsi, "LSTM based stock prediction using weighted and categorized financial news," *PLoS ONE*, vol. 18, no. 3, Mar. 2023, doi: 10.1371/journal.pone.0282234.

[50] Y. Zhang, Q. Liu, and L. Song, "Sentence-state LSTM for text representation," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Jan. 2018, pp. 317–327, doi: 10.18653/v1/p18-1030.

[51] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer, "The cross-entropy method for optimization," in *Handbook of Statistics*, 2013, pp. 35–59, doi: 10.1016/b978-0-444-53859-8.00003-5.

[52] K. Bian and R. Priyadarshi, "Machine learning optimization techniques: a survey, classification, challenges, and future research issues," *Archives of Computational Methods in Engineering*, Mar. 2024, doi: 10.1007/s11831-024-10110-w.

[53] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv:2010.16061*, Oct. 2020.

BIOGRAPHIES OF AUTHORS

Oussama Ndama is a Ph.D. student in data science, artificial intelligence, and smart systems research team (DSAI2S), C3S Laboratory, Faculty of Sciences and Technologies (FST), Tangier, Morocco. He had his Master's in Computer Science and Big Data, Laureate of FST of Tangier. He is also a Business Intelligence Engineer with more than 5 years of experience in different multinational companies. The research topics of interest are smart systems, machine learning, deep learning, NLP, ANN, sentiment analysis, and smart cities. He can be contacted at email: oussama.ndama@etu.uae.ac.ma.

Ismail Bensassi is a Ph.D. student in data science, artificial intelligence, and smart systems research team (DSAI2S), C3S Laboratory, Faculty of Sciences and Technologies (FST), Tangier, Morocco. He is an Engineer in computer Science, Laureate of FST of Tangier. The research topics of interest are smart connection of user profiles in a big data context, multi-agent systems (MAS), case based reasoning (CBR), ontology, machine learning, smart cities, and eLearning/MOOC/SPOC. He can be contacted at email: bensassi.ismail@gmail.com.

Safae Ndama is a Ph.D. student in data science, artificial intelligence, and smart systems research team (DSAI2S), C3S Laboratory, Faculty of Sciences and Technologies (FST), Tangier, Morocco. She earned her Master's in Computer Science and Big Data from the FST of Tangier. With two years of experience as a Data Scientist, her research interests include smart systems, data science, artificial intelligence, sentiment analysis, and smart cities. She can be contacted at email: safae.ndama@etu.uae.ac.ma.

Dr. El Mokhtar En-Naimi is a Full Professor in the University of Abdelmalek Essaâdi (UAE), Faculty of Sciences and Technologies of Tangier (FSTT), Department of Computer Sciences. He was temporary professor from 1999 to 2003 and permanent professor since 2003/2004 until now. He is a Full Professor in UAE, FST of Tangier. He was a head in Department of Computer Sciences, since October 2016 until the end of December 2020. He was responsible for a Licence of Science and Technology, LST Computer Engineering ("Licence LST-GI"), from January 2012 to October 2016. He is a chief of Data Science, Artificial Intelligence and Smart Systems (DSAI2S) Research Team since the academic year 2022/2023. He is also a founding member of the both laboratories: Laboratoire d'Informatique, Systèmes et Télécommunications (LIST) Laboratory (from 2008 to 2022) and Computer Science and Smart Systems (C3S) Laboratory since the academic year 2022/2023 until now, the University of Abdelmalek Essaâdi, FST of Tangier, Morocco. He is also an expert evaluator with the ANEAQ, since the academic year 2016/2017 until now, that an expert of the private establishments belonging to the territory of the UAE and also an expert of the initial or fundamental formations and formations continuous at the Ministry of Higher Education, Scientific Research and Executive Training and also at the UAE University and the FST Tangier since 2012/2013 until now. He is an author/co-authors of several articles, published in The International Journals in Computer Sciences, in particular, in multi-agent systems (MAS), cases based reasoning (CBR), artificial intelligent (AI), machine learning (ML), deep learning (DL), eLearning, MOOC/SPOC, big data, data-mining, wireless sensor network, VANet, MANet, and smart city. He was/is also director of several doctoral theses in computer sciences. He has too served as a general chair, technical program chair, technical program committee member, organizing committee member, session chair, and reviewer for many international conferences and workshops. In addition, he is an associate member of the ISCN-Institute of Complex Systems in Normandy, the University of the Havre, France, since 2009 until now. He can be contacted at email: en-naimi@uae.ac.ma.