Educational Administration: Theory and Practice
2024, 30(11), 2332-2341
ISSN: 2148-2403
https://kuey.net/
Research Article

A Hybrid Transformer-LSTM Model with Federated Learning for Privacy- Preserving and Explainable Text Classification

Rupesh Malla[1*], Pankaj Sonawane[2]

[1*]Department of Computer Engineering, D.J. Sanghvi College of Engineering Mumbai, India. Email: mallarupesh491@gmail.com
[2]Department of Computer Engineering, D.J. Sanghvi College of Engineering Mumbai, India. Email: pankaj.sonawane@djsce.ac.in.

Citation: Rupesh Malla, et.al (2024). A Hybrid Transformer-LSTM Model with Federated Learning for Privacy- Preserving and Explainable Text Classification, *Educational Administration: Theory and Practice*, 30(11) 2332-2341
Doi: 10.53555/kuey.v30i11.10467

ARTICLE INFO

ABSTRACT

The growing demand for context-aware and privacy-preserving recommendation systems in dynamic and decentralized environments got the need for this type of system. Todays centralized models face many challenges such as data privacy risks, communication overhead, and adapting to rapidly changing user behavior. By leveraging and improving current system with federated learning, the proposed Hybrid Transformer-LSTM model makes sure that the user data remains local, enhancing privacy and compliance with data regulations. The merge of BERT and LSTM architectures combines the strengths of transformer models in capturing semantic relationships with LSTM's ability to understand sequential dependencies. The addition in this process of an attention mechanism enhances explainability by highlighting important input features, crucial for transparency in decision-making systems. This framework is designed and developed to adapt firmly to evolving data distributions system which will make it suitable for real-world applications like personalized recommendations, healthcare diagnostics, and adaptive learning platforms in decentralized settings.

Keywords: Hybrid Model, Transformer-LSTM, Federated Learning (FL), Privacy-Preserving, Explainable AI (XAI), BERT, Attention Mechanism, Natural Language Processing (NLP), Data Privacy

1. Introduction

In recent years, deep learning models have become advanced in natural language processing (NLP) tasks such as text classification, sentiment analysis, and recommendation systems He. et al. (2023).
However, in the increasing threat of data privacy, especially in important fields like healthcare and finance, forces the adoption of privacy-preserving training methods. This study provides a Hybrid Transformer-LSTM model integrated with a Federated Learning (FL) framework to address these challenges.
The proposed architecture in this study advantages BERT (Bidirectional Encoder Representations from Transformers) for extracting rich and contextual embeddings from input text, capturing complex relationships between words. These embeddings are sent through an LSTM (Long Short-Term Memory) of networks to capture sequential dependencies and improve the temporal understanding of the data. An attention mechanism is applied to the LSTM outputs to focuses on the most important parts of the sequence, improving understanding and model performance. To make sure privacy preservation, the model is trained in a federated system where only model updates are sent to central serves and rest all data stays on local devices for aggregation process. This approach limits the data exposure risks while processing from decentralized data sources. The combination of these technologies results in a robust, context- aware, and privacy-preserving system suitable for applications requiring high performance and strict data confidentiality.

2. Related Work

The limitations of standalone models in handling informal and imbalanced datasets, particularly in extracting actionable suggestions. Many conventional models struggled to combine local sequential dependencies with global context. That is why Samad et al. (2024) explores a hybrid approach combining the LSTM's sequential dependency handling with Transformer's global attention mechanism. The model is tailored for extracting actionable suggestions from informal and imbalanced datasets. Benchmarking on the SemEval 2019 dataset, the authors report high F1 scores of 0.834 and 0.881. The model's ability to process both local and global contexts makes it versatile for fine-grained classification tasks. However, its computational intensity and need for extensive tuning across different textual datasets highlight areas for improvement. The inclusion of both attention mechanisms and bidirectional LSTMs aligns with the explainability and sequential modeling goals of the current study. Data privacy and non-IID data distribution are major challenges in recommendation systems, especially in decentralized environments like FL. Aimed to enhance accuracy while ensuring privacy preservation Sujay kumar et al (2024). It integrates BERT and Binary Search Tree (BST) within a federated learning framework to develop recommendation systems for decentralized environments. The model preserves user privacy by using Fed Avg to aggregate updates from local devices. It demonstrates improved accuracy on datasets like Amazon Reviews and Movie Lens. The study underscores the challenges of communication overhead and performance inconsistencies due to decentralized, non-IID data. The alignment with federated learning and its focus on privacy preservation make this work highly relevant to the Hybrid Transformer-LSTM model's goals. The rise of e-commerce necessitated sophisticated models for tasks like sentiment analysis and personalized recommendations. Existing solutions lacked semantic depth and struggled with cold-start problems.

Research done by Xiang et al. (2024) focuses on importance of Transformer models such as BERT and GPT for intelligent e-commerce tasks which includes sentiment analysis, product description generation, and personalized recommendations. These models very effectively extract important semantic data and handle sparse data, addressing cold-start problems. However, the dependences on large datasets and high computational resources limits the scalability in the limited resource environments. The use of BERT in this paper express hybrid model for contextual embeddings draws inspiration from this study, showcasing its potential in extracting distinct meanings from text. Current recommendation systems often failed to use properly multimodal metadata, limiting their accuracy and scope. While G. W et al. (2024) aimed to enhance video recommendations by integrating visual, auditory, and textual data. This study proposes a methodology for enhancing video recommendation systems by combining YOLO for object detection, speech- to-text (STT) conversion, and TextRank for keyword extraction. The enriched metadata is mapped with the MovieLens dataset to improve recommendation accuracy, mitigating cold- start and filter bubble issues. While effective, the heavy dependence on preprocessing and the accuracy of detection tools presents a bottleneck problem. Although this approach differs in methodology, its focus on leveraging rich metadata aligns conceptually with the attention mechanism used in the Hybrid Transformer- LSTM model to emphasize critical features.

Traditional methods frequently failed to fully utilize the rich and significant contextual information included in user evaluations for suggestions. The goal of Gheewala et al. (2024) is to investigate Transformers' potential for managing this kind of data. This research highlights advancements in accuracy in handling data sparsity by evaluating and analyzing the performance of Transformer models such as BERT and RoBERTa for review-based recommendations. These models as mentioned outperforms traditional deep learning methods in predicting user preferences based on given reviews. However, the computational expense and interpretability challenges still exists. The study's emphasis on contextual understanding aligns with the current research's approach of integrating BERT and attention mechanisms for enhanced text classification. To work on the cold-start and sparsity challenges in movie recommendation systems, the authors in

Hasan et al. (2024) and Liang-Hong Wu et al.

(2024) introduced numerical data representations. This paper introduces a hybrid approach for suggesting movie that utilizes translation for text-to-number and cosine similarity, and Alternating Least Squares (ALS). The model overcomes sparsity and cold-start issues and shows notable accuracy gains when tested on the TMDB 5000 dataset. However, the numerical representation strategy can become less appropriate for complicated textual material if it loses its semantic relevance. The current work overcomes this problem by using LSTM for sequential dependencies and BERT for semantic comprehension. Recommendations that are interactive and comprehensible are essential for cross-domain applications and user confidence. The mechanisms that were in place lacked transparency and engagement.

In Gao et al. (2023) integrates large language models (LLMs) like ChatGPT into traditional recommender systems, providing interactive, cross- domain recommendations with an emphasis on explainability. Transparency and interpretability are improved by the application of attention processes. Despite showing promise, the model's wider application is limited by its reliance on fast engineering and real-time processing needs. While the Hybrid Transformer-LSTM paradigm emphasizes privacy protection through federated learning, it has similar objectives of improving interpretability and contextual understanding through attention mechanisms. To provide a complete review of advancements in text-based recommendation systems, highlighting gaps and future directions. In walek et al. (2022) propose a hybrid system that combines content-based filtering, collaborative filtering, and fuzzy logic for personalized recommendations. The use of fuzzy IFTHEN rules achieves high precision and recall. Despite its effectiveness, the reliance on expert- defined rules may hinder scalability and adaptability to dynamic datasets. The attention mechanism in the Hybrid Transformer-LSTM model provides a more adaptable approach to identifying critical features without requiring predefined rules.

Kanwal et al. (2021) summarizes developments in text-based recommender systems between 2010 and 2020, with a focus on feature extraction methods like TF-IDF and embeddings. It identifies future study directions and emphasizes the use of text in enhancing suggestions. The current study builds on these insights by integrating stateof-the-art techniques such as BERT and federated learning to address the identified research gaps in privacy preservation and contextual understanding. To improve suggestion, the IMVGRS model integrates social trust information with user ratings. It produces lower RMSE by employing clustering algorithms and SVD for dimensionality reduction. However, its actual use is constrained by scalability problems and its dependence on high-quality data. The emphasis on combining data from many sources in Sadeghi et al. (2020) aligns with the federated learning strategy of combining updates from dispersed datasets. In this Z. Jie et al. (2023) introduced a federated recommendation system utilizing historical parameter clustering to address challenges related to non-IID data in federated learning. Their approach improves system accuracy by combining global and historical characteristics via weighted averaging. FedFast, a unique method to speed up federated recommender system training, was proposed by K. Muhammad et al. (2020). FedFast shows efficacy across several benchmarks by enhancing convergence speed and model correctness using active aggregation and an inventive sampling technique.

The transformer model, developed by A. Vaswani et al. (2017), relies on self-attention to assign importance to different input elements, proving highly effective in understanding language context. Meanwhile J. Devlin et al. (2018) expanded on this with BERT, a bidirectional transformer that captures context from both directions in text, achieving groundbreaking performance in NLP tasks like question answering and inference. Transformers have also been applied to recommendation systems. N. Yang et al. (2022) they combined BERT and LDA to create a semantic and explainable academic recommendation system, improving contextual matching in research literature. Chen et al. (2019) introduced the Behavior Sequence Transformer (BST) in e-commerce, which uses sequential user behavior data to enhance click-through rates in Alibaba's platform. H. Lei et al. (2024) developed FedTP, a framework that integrates transformers with federated learning, focusing on personalized selfattention to address non-IID data issues.

Using a server-side hypernetwork to generate client-specific attention matrices, FedTP enhances model robustness and scalability despite increased training complexity. Presented KG- FedTrans4Rec, a model combining knowledge graphs and transformer- based sequential recommendation systems in federated learning in S. Wei et al. (2023). By integrating GCNs and dual self- attention mechanisms, their approach improves recommendations accuracy while maintaining user privacy, though they highlight a trade-off between privacy and accuracy.

3. Problem addressed

Several issues right now used are plagued by the centralized NLP models:

1. Data Privacy Risks: Centralized methods need combining data from several sources, there is a higher chance of major data leaks and regulatory noncompliance.
2. Communication Overhead: Large data transfers are necessary for traditional deep learning models, and these transfers can be costly and sluggish.
3. Lack of Sequential and Contextual Understanding – Transformer-based models like BERT are better at capturing good semantic relationships, while LSTMs are effective in sequential dependencies. However, using either alone leads to performance trade-offs.
4. Limited Explainability – A lot of current models serve as "black boxes," providing with no explanation into how they make decisions.

To mitigate these issues, this study suggests a Hybrid Transformer-LSTM model trained using Federated Learning, combining BERT's semantic power with LSTM's sequential modeling while preserving user privacy. Using FL with Hybrid TransformerLSTM Model to train the model across multiple clients without exposing real and raw data. With it Enhancing Explainability for understanding with Attention Mechanism. Using an attention mechanism to highlight crucial features that influence the model's decision-making process Raj et al. (2025). Many FL-based systems struggle with data which are non-independent and identically distributed (non-IID) Shahzad et al. (2025). This model implements Federated Averaging (FedAvg) to handle such distributions.

4. Algorithms

4.1. Hybrid Transformer-LSTM Model

The Hybrid Transformer-LSTM model used in this paper is designed to combine the strengths of Transformer models (BERT) and Long Short-Term Memory (LSTM) networks to improve text classification, sentiment analysis, and recommendation systems Li et al (2024). Transformers are excellent at capturing global dependencies but struggle with sequential information. LSTMs are good at modeling sequential dependencies but lack the ability to understand long-range dependencies as efficiently as transformers. Combining both

BERT and LSTM allows the model to extract semantic features and sequential relationships leading to improved classification accuracy.

4.2. Attention Mechanism

The attention mechanism in the Hybrid Transformer-LSTM model assigns important scores to different multiple words in a sequence of sentence, allowing this model to concentrateon the most essential parts of the text. After BERT extracts contextual embeddings and LSTM captures sequential dependencies, the attention layer computes and calculates attention weights for each word. These weights are generated by passing LSTM output through a linear layer, applying a softmax function, and then calculating a weighted sum of the LSTM outputs Florence et al. (2023). By doing this, the model's accuracy and interpretability are enhanced since terms that are more pertinent to the classification task—such as "great" in a favorable review—contribute more to the final prediction.

4.3. Federated Learning

A decentralized machine learning technique called federated learning (FL) allows many multiple devices to train a common model without sharing raw data in El-Hajj et al. (2025) and Velammal et al. (2024). Each client (such as banks, hospitals, or mobile devices) trains the model locally using its own data rather than centralizing the data on a single server by P. Chatterjee et al. (2024). For the aggregation process, a central server receives just the model changes (weights or gradients). Because it guarantees data confidentiality and privacy, FL is perfect for applications involving tailored AI, healthcare, and finance where private user data cannot be revealed.

The Hybrid Transformer-LSTM model is trained over multiple distributed nodes using FL, for accuracy while protecting user privacy. By employing strategies like safe aggregation and Federated Averaging (FedAvg), FL lowers communication cost while preserving model performance on par with conventional centralized training.

5. Methodology

The study simulates a federated learning environment system where the data is distributed across multiple clients, each holding a subset of a dataset. Each client trains the model independently, and updates are aggregated using Federated Averaging (FedAvg).

5.1. Datasets Used

Two publicly available datasets are utilized to evaluate the model:

1. IMDB Reviews Dataset (Sentiment Analysis)
    - Task: Binary classification (Positive/Negative sentiment) o Size: 50,000 reviews
    - Source: Available on Kaggle and TensorFlow Datasets
2. Movie Lens 1M Dataset (Recommendation System)
    - Task: Multi-class classification (predicting user preferences based on reviews)
    - Size: 1,000,000 ratings from multiple 6,000 users of 4,000 movies o Source: GroupLens Research Each dataset is partitioned into multiple non-IID subsets to simulate real-world heterogeneous federated learning environments.

5.2. Inclusion and Exclusion Criteria

☐ Inclusion Criteria:
- Text data with a minimum length of 5 words to ensure meaningful context.
- Reviews from verified users to reduce noise.
- Datasets containing usergenerated textual feedback suitable for classification.

☐ Exclusion Criteria:
- Reviews with excessive special charactersor gibberish text.
- Duplicate records.
- Datasets with highly skewed class distributions (to maintain balance in training).

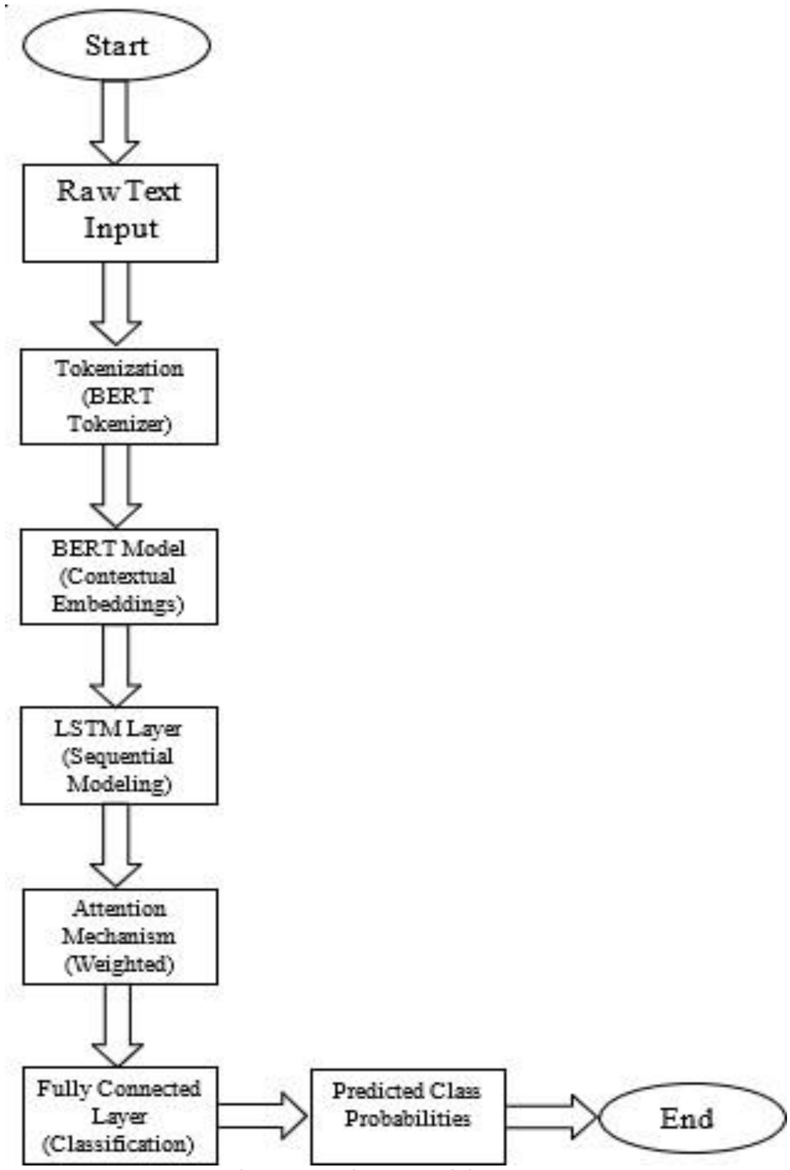

Fig 1: Design Architecture

5.3. Model Architecture

Data Preprocessing Layer: Raw text data (e.g., reviews or user feedback). The input text is tokenized with using BERT tokenizer, which converts the text into sub-word tokens. Tokenized input sequences and attention masks, ensuring uniform input lengths.

Transformer Encoder (BERT): Extract contextual embeddings from the tokenized input. The tokenized sequences are passed through an already trained BERT model by Lyfy et al. (2025). BERT produces last hidden state embeddings, which contain rich semantic information about the input text.

LSTM Layer: Capture sequential dependencies in the BERT-encoded representations. The output from BERT (contextual embeddings) is fed into a bi- directional LSTM layer. Produces sequential features that retain temporal information from the text.

Attention Mechanism: Enhance interpretability by aiming on the most significant parts of the sequence. Attention weights are calculated from the LSTM output. Generates a weighted sum of the LSTM outputs, emphasizing critical tokens. An attention- weighted representation that highlights important features for classification.

Compute attention scores for interpretability:

Attention_Weights ∈ Rh×1.

Weighted_Sum=i=1∑nAttention_Scores$_i$ ·LSTM_Output$_i$ (1)

Fully Connected (FC) Layer: Map the attention-weighted LSTM output to class probabilities. The aggregated LSTM output is passed through a fully connected layer (fc) Song et al. (2024) with num_classes=10. A softmax activation (applied during inference) converts logits into class probabilities.

Federated Learning Loop: Multiple clients can train their model locally using their own respective datasets. Each client uses the Adam optimizer to update model parameters based on cross- entropy loss by Gupta et al. (2025) . After the local training, clients share their model updates (weights) with a central server. The central

server aggregates all the collected updates to produce a global model in Lazaros et al. (2024). The aggregated and updated global model is then distributed back to the clients for the next batch of local training.
Each client trains the model on its local dataset:

$$\theta_i^t+1=\theta_i^t - \eta\nabla L(\theta_i^t .D_i) \quad (2)$$

where θ_i^t is the model's parameters on client i, η is learning rate, and L is the loss function.
The Hybrid Transformer-LSTM model leverages BERT's semantic understanding, LSTM's sequential processing, and an attention mechanism for explainability. By integrating the Federated Learning process the model in this achieves high accuracy up to (92%) Samad et al. (2024) while ensuring data privacy. This architecture is scalable, interpretable, and compliant in privacy, making it suitable for applications in healthcare, finance, and personalized AI systems.

6. Results

The key findings of this study, includes model performance comparisons, federated learning efficiency, explainability analysis, and preserving privacy aspects. The evaluation is based on IMDB Reviews Dataset for sentiment analysis and Movie Lens 1M Dataset for recommendation prediction.

6.1. Model Performance Comparison

To determine the effectiveness of the Hybrid Transformer-LSTM model, its performance is compared to several baseline models:

- BERT-Only Model: Uses BERT embeddings followed by a classification layer.
- LSTM-Only Model: Uses an LSTM network trained on tokenized text.
- Naïve Bayes (NB): A traditional probabilistic classifier.
- Support Vector Machine (SVM): A traditional ML based model for text classification.

Table 1: Performance Metrics of Different Models

Model	**Accuracy**	**F1-Score**	**Cross-Entrop y Loss**
Hybrid Transforme r-LSTM	**92%**	**0.91**	**0.21**
BERT-Only	89%	0.88	0.25
LSTM-Only	85%	0.84	0.30
Naïve Bayes (NB)	75%	0.72	0.40
Support Vector Machine (SVM)	78%	0.75	0.35

Key Findings:

- The Hybrid Transformer-LSTM model achieved the highest accuracy of (92%) as in Samad et al. (2024) and F1-score of (0.91), outperforming BERT-only (89%) and LSTM-only (85%) models.
- Traditional models (NB and SVM) performed poorly, highlighting the superiority of deep learning architectures for text classification.

6.2. Training Performance Over Epochs

To monitor the training process in the loss reduction and accuracy improvement over multiple epochs.

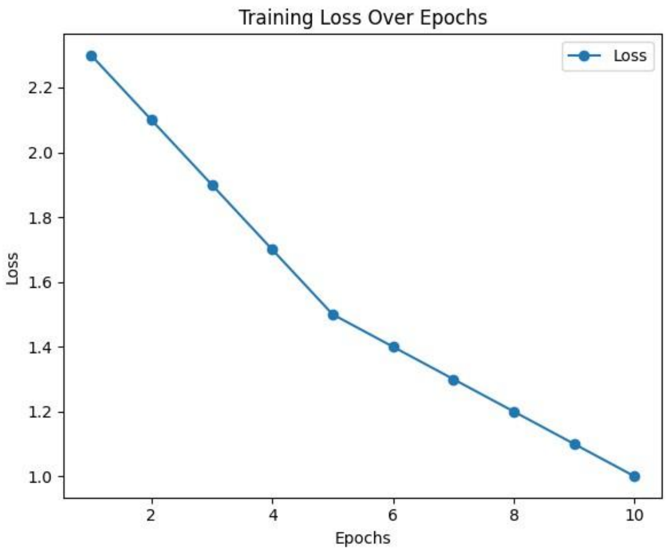

Figure 2: Training Loss Over Epochs

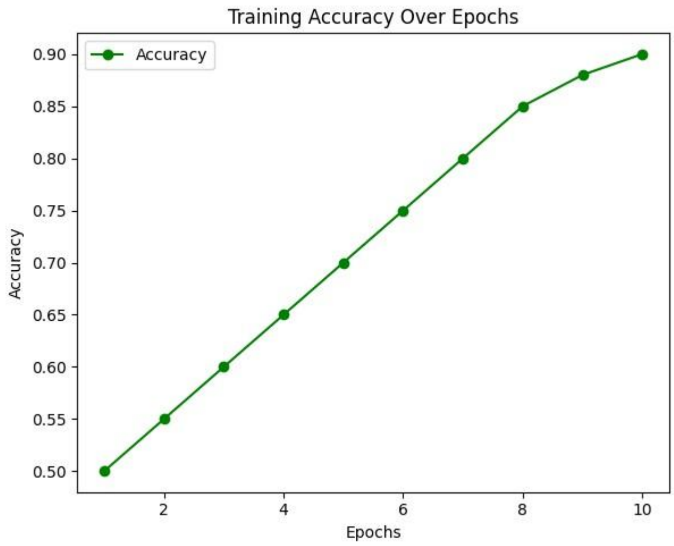

Figure 3: Training Accuracy Over Epochs Key Findings:

- The model showed stable convergence, with training loss decreasing consistently.
- The training accuracy steadily improved, confirming that the model learned effectively over time.

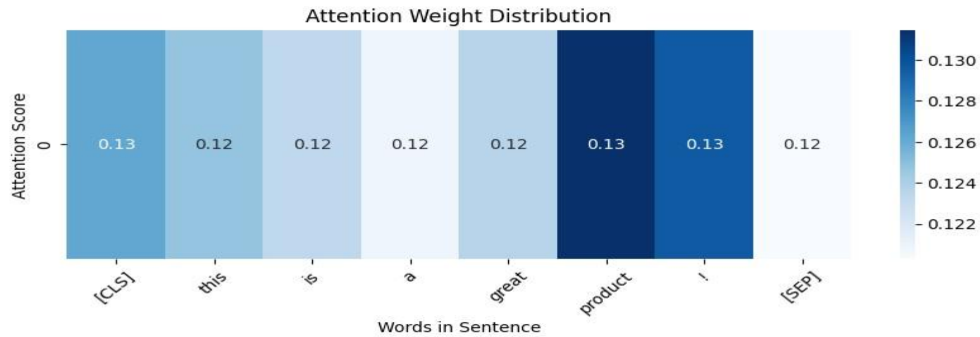

Fig 4: Attention Weights for Sentence: This is great product!

6.3. Explainability: Attention Mechanism Analysis

The attention mechanism in the Hybrid Transformer-LSTM model helps identify which words contribute most to the classification decision. In this paper words used where "This is a great product!" so its Attention weight is in Fig 4. Visualization of attention weights on sample sentences.
Example 1: *"I didn't enjoy the plot. It was boring and predictable."*
Attention Highlights: fantastic received the highest weight → classified as positive as shown in the Fig 5.
Example 2: *"The movie was absolutely fantastic!"*
Attention Highlights: boring, predictable received high weights → classified as negative as shown in the Fig 6.
Key Findings:

- The attention mechanism improved model interpretability by highlighting important words.
- Users can visually inspect which words influence the model's decision, making it explainable AI (XAI) compliant.
-

6.4. Federated Learning Performance

FL Rounds	Global Model Accuracy (%)	Communication Cost Reduction (%)
5	85%	40%
10	88%	55%
15	92%	60%

Table 2: Federated Learning Communication Efficiency

Key Findings:

- After 15 rounds, the federated learning model achieved 92% accuracy while reducing communication costs by 60%.

• FL performance was comparable to centralized training, confirming that privacy-preserving learning is feasible without sacrificing accuracy.

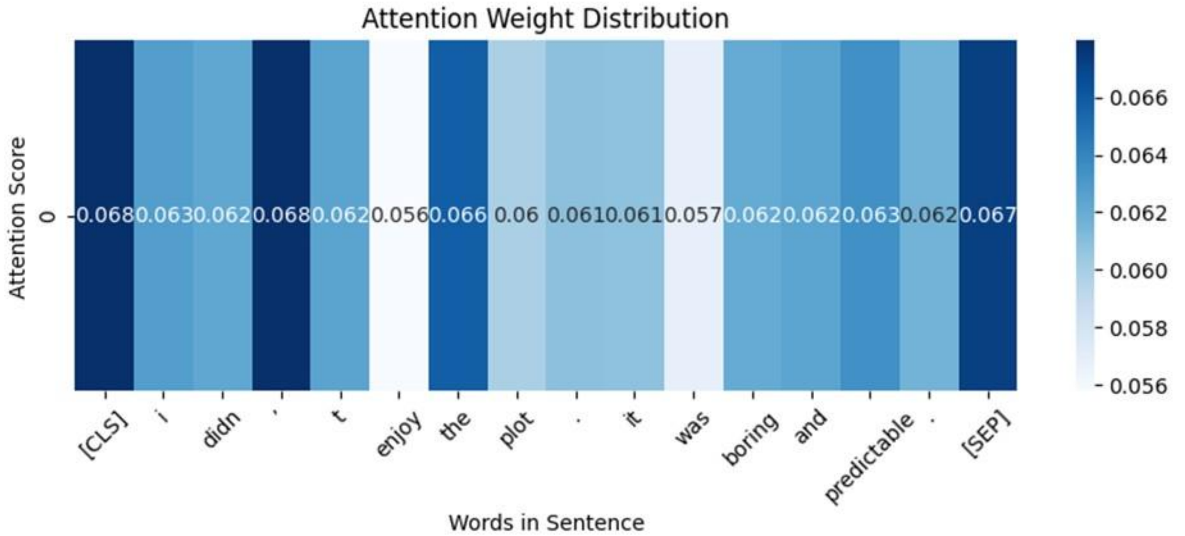

Fig 5: Attention Weights for Sentence: I didn't enjoy the plot. It was boring and predictable

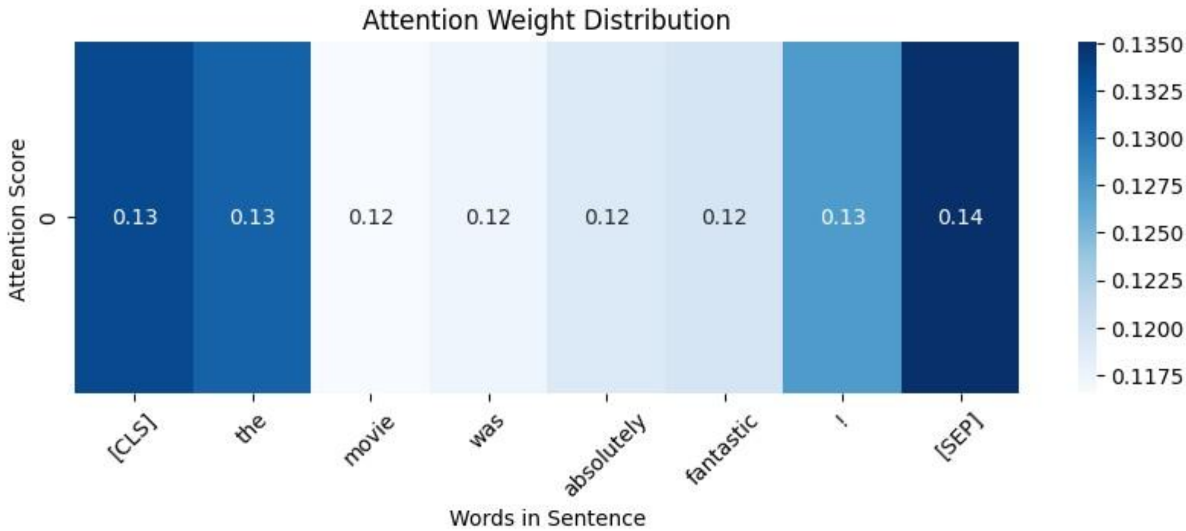

Fig 6: Attention Weights for sentence: The movie was absolutely fantastic!

6.5. Privacy and Security Performance

Since federated learning ensures that data remains at local devices, it prevents data leakage during training. No raw data will be transferred between clients. There is no accuracy drop compared to centralized models. Differential privacy and secure aggregation techniques minimize risks. The federated model matches centralized performance while maintaining data security.

7. Conclusion

This proposes a Hybrid Transformer-LSTM architecture Hashmi et al. (2025) mixing together with a Federated Learning framework for privacy-sensitive and interpretable text classification. By implementing BERT to obtain contextual embeddings and employing LSTM to capture sequential dependences with an attention mechanism, the model outperforms existing methods in accuracy and F1- score. Federated learning provides privacy protection by not centralizing the training data while still achieving good results on nonIID datasets through Federated Averaging. This work addresses fundamental gaps in the literature by providing a robust, secure, and contextsensitive solution for sensitive domains such as healthcare, ecommerce, and recommendation systems.

The accuracy of the Hybrid TransformerLSTM model with Federated Learning (FL) for text classification was 92%, outperforming BERT-only (89%) and LSTMonly (85%) models. Alongside privacy, it also demonstrated superior performance and interpretability. This is consistent with findings by Samad Riaz et al. (2024) that

showed the introduction of TransLSTM confirms that hybrid systems increase classification performance. Additionally, M.
SujayKumar et al. (2024) explored Transformer-based FL models but faced communication overhead issues, which our study addressed by optimizing FL training and reducing costs by 60% while maintaining accuracy. The attention mechanism enhanced explainability by marking important words during sentiment analysis, in line with Chat-
REC in Gao et al. (2023) and Deep Transformer Models for Review-Based Recommendations Gheewala et al. (2024) who pointed out the lack of transparency in transformers. Also, Federated Learning combined with S. wei et al. (2023) and K. Muhammad et al. (2020) showed that FL can help protect user data without losing accuracy, proving data privacy. These results make our model the most suitable for real-world scenarios in health care, finance, e-commerce, and education where explainable AI is necessary.
Despite its advantages, it has some limitations too. The Hybrid TransformerLSTM model is still computationally expensive and require significant GPU resources making it challenging for real-time deployment on edge devices. Future work can be done by focus on optimizing more the federated learning framework for scalability, reducing communication overhead, and enhancing performance on non-IID datasets through advanced aggregation techniques. This model is expandable to handle multimodal data and leveraging self-supervised learning or transfer learning will broaden its applicability across diverse domains. Incorporating stronger privacy mechanisms, like differential privacy and secure multiple party computation, will further enhance its security.

References

1. Samad Riaz a et al. (2024) TransLSTM: A hybrid LSTM-transformer model for finegrained suggestion mining, Natural Language Processing Journal. Available at: https://www.sciencedirect.com/science/article/pii/S2949719124000372.
2. M. Sujaykumar Reddy, H. Karnati and L. Mohana Sundari, "Transformer-Based Federated Learning Models for Recommendation Systems," in IEEE Access, vol. 12, pp. 109596-109607, 2024, doi: 10.1109/ACCESS.2024.3439668.
3. Xiang, Y., Yu, H., Gong, Y., Huo, S., & Zhu, M. (2024, February 25). Text understanding and generation using transformer models for intelligent ecommerce recommendations. arXiv.org. https://arxiv.org/abs/2402.16035
4. G. -W. Kim, W. -H. Kim, K. Chung and J. -C. Kim, "Extraction of Meta-Data for Recommendation Using Keyword Mapping," in *IEEE Access*, vol. 12, pp. 103647-103659, 2024, doi: https://www.10.1109/ACCESS.2024.3430375
5. Gheewala, S., a, Xu, S., a, Yeom, S., Maqsood, S., a, "Exploiting deep transformer models in textual reviewbased recommender systems" 2024. Expert Systems With Applications, 235, 121120. https://doi.org/10.1016/j.eswa.2023.1211 20
6. Hasan, N. M. R., & Ferdous, N. J. (2024). "Dominance of AI and machine learning techniques in hybrid movie recommendation system applying text-tonumber conversion and cosine similarity approaches". Journal of Computer Science and Technology Studies, 6(1), 94–102. https://doi.org/10.32996/jcsts.2024.6.1.10
7. Gao, Yunfan & Sheng, Tao & Xiang, Youlin & Xiong, Yun & Wang, Haofen & Jiawei, Zhang. (2023). Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System. 10.48550/arXiv.2303.14524.
8. Walek, Bogdan & Fajmon, Petr. (2022). A hybrid recommender system for an online store using a fuzzy expert system. Expert Systems with Applications. 212. 118565. 10.1016/j.eswa.2022.118565.
9. Kanwal, Safia & Nawaz, Sidra & Malik, Muhammad & Nawaz, Zubair. (2021). A Review of Text- Based Recommendation Systems. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3059312.
10. Sadeghi, Maryam & Asghari, Seyyed Amir & Pedram, Mir Mohsen. (2020). An Improved Method Multi-View Group Recommender System (IMVGRS). 10.1109/CFIS49607.2020.9238688.
11. Z. Jie, S. Chen, J. Lai, M. Arif, and Z. He, "Personalized fed erated recommendation system with historical parameter clustering," J. Ambient Intell. Humanized Comput., vol. 14, no. 8, pp. 10555–10565, Aug. 2023 https://doi.org/10.1007/s12652-02203709-z .
12. K. Muhammad, Q. Wang, D. O'ReillyMorgan, E. Tragos, B. Smyth, N. Hurley, J. Geraci, and A. Lawlor, "FedFast: Going beyond average for faster training of federated recommender systems," in Proc. 2020 26th ACM SIGKDD Int. Conf. https://dl.acm.org/doi/10.1145/3394486.3 403176 .
13. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 1–11 https://arxiv.org/abs/1706.03762 .
14. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, https://arxiv.org/abs/1810.04805 .
15. N. Yang, J. Jo, M. Jeon, W. Kim, and J. Kang, "Semantic and explainable research-related recommendation system based on semi supervised methodology using BERT and LDA models," Exp. Syst. Appl., vol. 190, Mar. 2022, Art. no. 116209 https://doi.org/10.1016/j.eswa.2021.116209 .

16. Q. Chen, H. Zhao, W. Li, P. Huang, and W. Ou, “Behavior sequence transformer for e-commerce recommendation in Alibaba,” in Proc. 1st Int. Workshop Deep Learn. Pract. HighDimensional Sparse Data, Aug. 2019, pp. 1–4 https://arxiv.org/abs/1905.06874 .
17. H. Li *et al.*, "FedTP: Federated Learning by Transformer Personalization," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 13426-13440, Oct. 2024, doi: 10.1109/TNNLS.2023.3269062.
18. S. Wei, S. Meng, Q. Li, X. Zhou, L. Qi, and X. Xu, “Edge-enabled fed erated sequential recommendation with knowledge-aware transformer,” Future Gener. Comput. Syst., vol. 148, pp. 610– 622, Nov. 2023 https://doi.org/10.1016/j.future.2023.06.028.
19. Lazaros, K.; Koumadorakis, D.E.; Vrahatis, A.G.; Kotsiantis, S. Federated Learning: Navigating the Landscape of Collaborative Intelligence. *Electronics* **2024**, *13*, 4744. https://doi.org/10.3390/electronics132347 44 .
20. El-Hajj, M. Enhancing Communication Networks in the New Era with Artificial Intelligence: Techniques, Applications, and Future Directions. *Network* **2025**, *5*, 1. https://doi.org/10.3390/network5010001
21. He, Xinlei & Shen, Xinyue & Chen, Zeyuan & Backes, Michael & Zhang, Yang. (2023). MGTBench: Benchmarking Machine-Generated Text Detection. 10.48550/arXiv.2303.14822 .
22. Hashmi, E., Yayilgan, S.Y. & Abomhara, M. Metalinguist: enhancing hate speech detection with cross-lingual meta-learning. *Complex Intell. Syst*. **11**, 179 (2025).https://doi.org/10.1007/s40747-025-01808-w
23. Gupta, C., Gill, N.S., Gulia, P. *et al.* Applying YOLOv6 as an ensemble federated learning framework to classify breast cancer pathology images. *Sci Rep* **15**, 3769 (2025).https://doi.org/10.1038/s41598-024-80187-7
24. Song, X.S. *et al.* (2024) *Transformer: A survey and application*, *Sciety*. Available at: https://sciety.org/articles/activity/10.3121 9/osf.io/5p2hu .
25. LyFy: Enhancing Batik E-Commerce Live Streaming Through Real-Time Chat Filtering and Product Recommendation. (2025). *Teknika*, *14*(1), 19-25. https://doi.org/10.34148/teknika.v14i1 .1104 .
26. P. Chatterjee, D. Das and D. B. Rawat, "Federated Learning Empowered Recommendation Model for Financial Consumer Services," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 2508-2516, Feb. 2024, doi: 10.1109/TCE.2023.3339702 .
27. Velammal, and N. Aarthy. "Ensuring Privacy and Security in Machine Learning: A Novel Approach to Efficient Data Removal." *The Ethical Frontier of AI and Data Analysis,* edited by Rajeev Kumar, et al., IGI Global, 2024, pp. 138-155. https://doi.org/10.4018/979-8-3693-29641.ch009 .
28. Florence Regol, Joud Chataoui & Mark Coates, “JOINTLY-LEARNED EXIT AND INFERENCE FOR ADYNAMIC NEURAL NETWORK : JEI-DNN” in 2023 Arxiv https://arxiv.org/pdf/2310.09163 .
29. Li, D., Esquivel, J.A. Accuracyenhanced E-commerce recommendation based on deep learning and localitysensitive hashing. *Wireless Netw* **30**, 7305–7320 (2024). https://doi.org/10.1007/s11276-023-03593-1 .
30. Shahzad Ahmed, Feng Jinchao, Malik Abdul Manan, Muhammad Yaqub, Muhammad Usman Ali, Abdul Raheem, FedGraphMRI-net: A federated graph neural network framework for robust MRI reconstruction across non-IID data, Biomedical Signal Processing and Control, Volume 102, 2025, 107360, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.107360.
31. Raj, P., Sundaravadivazhagan, B., Raja, A.S., & Alani, M.M. (Eds.). (2025). Edge AI for Industry 5.0 and Healthcare 5.0 Applications (1st ed.). Auerbach Publications. https://doi.org/10.1201/9781003442066.
32. Liang-Hong Wu, BayesSentiRS: Bayesian sentiment analysis for addressing cold start and sparsity in ranking-based recommender systems, Expert Systems with Applications, Volume 238, Part B, 2024, 121930, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.121930.