

DEEP LEARNING-BASED CREDIT RISK PREDICTION USING SELF-ATTENTION DRIVEN HYBRID LSTM MODEL

Dr. N. Penchalaiah, M.Tech, Ph.D¹, Devarala Divya²

¹Professor, Department of CSE, Audisankara College of Engineering and Technology, Gudur, India.

Email: drneeli.cse@audisankara.ac.in

²M.Tech Student, Department of CSE, Audisankara College of Engineering and Technology, Gudur, India.

Email: devarala.divya2002@gmail.com

Received: 25th May, 2026; **Revised:** 6th June, 2026; **Accepted:** 8th June, 2026; **Available Online:** 09th June, 2026

ABSTRACT

In order to mitigate the shortcomings of deep learning and machine learning, this extended paper offers a Self-Attention based Hybrid LSTM credit score prediction model. In order to have the most contextual information and minimize the amount of irrelevant patterns, the network emphasizes self-attention that concentrates on purchasing the most significant temporal and behavioral variables on high-dimensional financial data. The enhancement of data quality and stability is achieved by such techniques as temporal modeling, Min-Max normalization, to balance the data, SMOTE to select the features and PCA to reduce the dimensionality.

The technique is feasible because a financial company can input data file containing customer details and get real-time information on risk assessment to forecast a credit score through an internet application on Flask. Based on the experimental findings, the Self-Attention Hybrid LSTM is better than the base models in their accuracy, data generalisation, and financial power. This is the improved process of assessing the credit risk which has become intelligent, scalable and efficient.

Keywords: Feature Selection, PCA, SMOTE, Deep Learning, Sequential Data Modeling, Flask Deployment, Real-Time Credit Prediction, Financial Data Analytics, Self-Attention, Hybrid LSTM, Credit Scoring Prediction, Financial Risk Assessment, Temporal Feature Learning, Imbalanced Data.

How to cite this article: Penchalaiah N, Divya D. Deep Learning-Based Credit Risk Prediction Using Self-Attention Driven Hybrid LSTM Model. *Int J Drug Deliv Technol.* 2026;16(57s): 1486-1491. DOI: 10.25258/ijddt.16.57s.151

Source of support: Nil.

Conflict of interest: None.

I. INTRODUCTION

There are also the high-dimensional and behaviour-based databases that the digital financial services are quickly combined to forecast the credit scores. The universal machine learning practices cannot limit dynamically moving financial patterns and time-strict user conduct owing to continuous characteristics. A large amount of more recent deep and prompt based model literature has studied severe contextual representation with a wide variety of inputs, especially on small, noisy, or heterogeneous financial input. The information combination and selective focusing stimulates the new designs as per feature-oriented modelling and representation learning so that understanding of the models can be made easy.

It is further emphasized that the need to have model flexibility and informed representation learning in intricate classification plans is justified by the notion that few-shot and prompt-driven learning algorithms have the potential to reach a significant amount of contextual data with limited supervision. Contrary to the claims by Schick and Schutz [1] and Hu et al. [2] that the knowledge structure and cloze-style prompts resulted in enhanced classification strength, the article of Sun et al. [3] also revealed that giant language models are also able to generalize in categorizing situations. Han et al. [4] applied rule-guided quick modification to increase decision boundaries but Wen and Fang [5] applied

graph-grounded pre-training to increase the performance of low-resource classification.

As a result of these improvements, it has now deployed an aspect of self-attention in the Hybrid LSTM model to recognize the proper kind of behavioral and temporal properties of the credit scoring data. To enhance sequential awareness, the model establishes long-term relationships and abnormality of features and attends much to the concerned financial trends. It is a combination of the RFE and the PCA which identifies the few predictors of interest as well as SMOTE in order to balance the very skewed credit data with the aim of enhancing the reliability and generalization of the models.

Flask-based web application is the real life application of the improved model to predict credit rating in real time, through the use of financial data files. This relationship will give interpretable and scalable and valid base to the existing financial institutions that achieve high data setting change rates. The strategy uses the existing intelligent feature learning and flexible deep neural structure to incorporate intelligent information to the conventional credit scoring strategy.

II. LITERATURE SURVEY

The modern processes of text classification will misclassify the adversarial phenomena which are perverted versions of the original text which the human beings will have a hard time detecting. The current NLP researches are

based on the substitution of the hostile cases on the basis of the rule of synonym. There are ways of creating the token changes which may be overly complex and irrelevant and can be single-handedly identified by the individuals. BAE is a type of black box attack, and it perturbs a BERT masked language model by means of context to produce adversarial samples. BAE alters and appends the tokens by covering one of the parts of the original text and comprises the alternative through the usage of BERT-MLM. We show that BAE generates a stronger attack and generates adversarial examples which are more grammatical and semantically coherent on both automatic and human evaluation systems [6].

The deep learning models perform better than machine learning-based models in sentiment analysis, question answering and natural language inference, and news classification. We contrast more than 150 text categorization models that are now accessible in deep learning, its application, technological benefits, and the commonalities. We are also about 40 in our consideration of popular text categorization data. Finally, but not the least, we provide the quantitative analysis of the performance of the deep learning models on the most popular benchmarks and the future research [7].

MixText, a semi-supervised text categorization learning algorithm, is presented in this paper which is based on the recent technology of TMix data augmentation. TMix offers series of enhanced training situations through interpolation of training materials in concealed area. At this stage, the new trends in data augmentation make it possible to draw low-entropy labels on data that has no labels and is thus as easy to implement as labeled data. MixText has also competed with all the text classification benchmarks than the pre-trained and refined models on the top of the state-of-the-art semi-supervised training systems. Supervision is absent and improvement is also actualized [8].

The aqueous zinc (Zn) metal batteries have great prevalence as these are inexpensive and harmless, however due to the advancement of dendrites, hydrogen evolution reaction and corrosion reaction, the aqueous zinc metal batteries have restricted their application. Cyclen organic molecule has short molecules which is beneficial to Zinc sulfate electrolyte (1.4.7.10). The deposition of the Zn^{2+} is uniformly deposited because the deposition is made after the Zn (002) plane. It is more structured as compared to H_2O molecules since it is cyclic in nature and contains N. This enabled the Zn||Zn cell to recycle at 1 mA cm² and even 1 mAh cm² with an approximated value of 2400 hours. Addition of cyclen to Zn||V2O5 full cells also enhanced the retention capacity of the cell at 90.9% level at 3 A g⁻¹ at the 2000 cycles [9].

An idea that we introduce is a noisy channel approach to applying language models to few-shot text classification to prompt. Each word of the input should be described by channel models because they compute the conditional probability of the input and not the probability of the label given the input (direct models). Recently spawned few-shot

learning technologies are trained on our channel models, and need in-context presentation or quick switching of change parameters of language models. We have discovered that channel models are superior to direct models in both instances as they are more stable (have less variance and worst-case accuracy). We also possess the so-called complete ablations which would be equivalent to channel prompt tuning rather than direct head tuning where the labels of the training data are disproportionate, there is no instance of the training data or the training data is generalized to unknown labels [10].

III. METHODOLOGY

The method proposed can be used to elucidate the time-related relations and explain the effectual factors of the behavioral patterns of the credit rating samples based on the enhanced Hybrid LSTM model and self-attention framework. Most useful predictors sorted by RFE are the most useful, PCA simplifies the learning process by reducing the dimension and SMOTE equalizes the classes. Min-Max is used to scale the numerical features. The LSTM layers are fed with more superior data to support learning longer time patterns. The traditional models do not have contextual information and this information is enhanced by a self attention layer which attempts to quantify the useful time steps in an adaptative manner. On the binary cross-entropy and optimization constraints, Hybrid is more expected to transform and differentiate between good and bad accounts of credit in the long-term. Having trained and tested the model on a sample financial dataset, the final prediction system may be deployed using a Flask-stable web interface to be in a position to evaluate the credit scores in real-time using the input user data. It is a highly practical, interpretable and scalable model, which would be applied by the contemporary financial institutions who need precision in respect of predicting credit risks.

A. Proposed Undertaking:

The paper develops a deep learning algorithm which entails time modelling, taking into account the scenario, and intelligent data preparation of enhanced forecastability of credit scores. The traditional credit rating does not provide multidimensional data and multifaceted relationships in dynamic financial trend. The model has a hybrid LSTM structure to distinguish between minor patterns of sequential data and major financial events, which have an attention system. The system can satisfy the complexity of dependencies, which include the spending cycles, alteration of repayment and the behavior change, which is correlated with risk and increase the precision and dependability of the forecasts. Min-Max normalization, SMOTE oversampling, dimensionality optimization of RFE, dimensionality optimization of PCA provide good and equal learning and data used in the modeling.

The implementation of Flask to facilitate the real-time decision-making tool based on the proposed model is also the objective of the current project. Through the hybrid deep learning, financial companies can estimate statistics on credit risk by submitting channeled datasets of structured or

semi-structured information on our platform of deployment. The modern lending is able to utilize the calculational efficiency, scale and usability of the technology in the lending. The novelty of the research within the limits of the present program is complemented by its application in the framework of the application of deep learning in credit risks assessment process. This is as a result of the interpretation intelligence, deployable, intelligent credit rating instruments that facilitate the enhanced practical decision making in the financial sector.

B. System Architecture:

Multi-spatial analysis of financial data in the shape of attention-based augmentation and sequential modelling can be achieved with the help of the deep learning model with several layers. The ability to control imbalance through Min-Max normalization with the help of SMOTE, the extraction of RFE features and dimensionality reduction via PCA can all be considered as being part of the initial process of the data preparation. The Hybrid LSTM layer is used to learn the time dependent relationship of financial records of the users on the basis of optimized and purified dataset. Their most important are dynamically weighted time steps and behavioral patterns to the LSTM layers, and the self-attention mechanism is indirectly opened to the surface, where financial indicators of payback cycles, expenses abnormalities and risky sequences are presented. It is also related closely with architecture as well as hybrid lossing functionality that is applied in order to implement homogeneous learning and categorization of multifaceted financial image.

Evaluation of prediction modules and communication with the users through deployment in Flask. The results of online interface can be displayed on estimates of the model credit scores using CSV or Excel files of user data provided by financial analysts or financial institutions. Its back-end contains an API layer that gives the feature of risk-classification of the prediction engine. The resultant deployment design will have the attributes of scalability; low monetary construction incorporation; and faster completion of massive datasets. The design will entail deep learning in terms of strong elements and web delivery system to develop a consistent and understandable credit scoring mechanism and a production ready one.

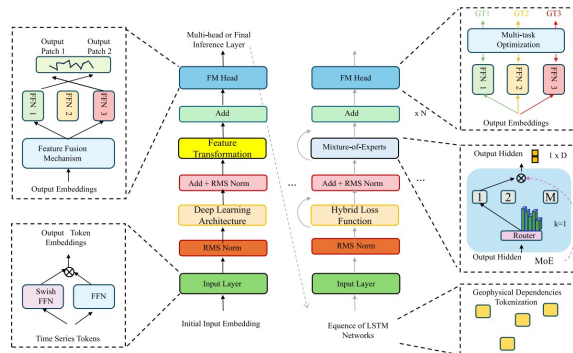


Fig. 1. Proposed architecture

IV. IMPLEMENTATION

1. MODULES:

a. *Data Preprocessing Module*

Every adjustment of data that would be necessary to create an effective learning of the model is done in this module. The data goes through the PCA to reduce the size of the dimensions and retain the useful data, the SMOTE is applied to the data to eliminate the disproportion of classes, the RFE is applied to the data to deliver the most predictive variables and the Min-Max normalization of the numerical variables is applied. This ensures availability of utopian, well balanced and sanitary information that is availed to the deep learning pipeline.

b. *Feature Engineering & Extraction Module*

The type relies on structural and behavioral financial variables to come up with time-related differences in credit rating. It transforms history of transactions, history of repayment and spending patterns into deep learning time-series representations, which enhance raw data, and sorts the feeds to LSTM model.

c. *Hybrid LSTM Modeling Module*

A major module is a long term dependency determination, behavior change and pattern as much as risks are involved in user history data according to sequence financial data and stacked LSTMs. It simplifies prediction model, it transforms time sequence to meaningful internal representations.

d. *Self-Attention Mechanism Module*

The adaptive weights are applicable in this case to improve the work of LSTM when time rates are large or monetary events are involved. The overruling items of data are also the high impact behaviors of the late payments and unexpected expenses. This is one of the significant predictors and situational data.

e. *Classification & Hybrid Loss Module*

This module hedges the binary credit risk (good vs. poor) using the assistance of fully correlated layers. It solves the challenge of the discrimination of classes and its convergence and is augmented with the assistance of a hybrid loss function i.e., the binary cross-entropy and optimization constraints are taken into account.

f. *Model Training & Validation Module*

Through this module, all the data amendments required to obtain the model are attained. Reduction of the dimensions can be achieved with the assistance of PCA and a significant amount of data can be stored, the asymmetry between the classes can be eliminated with the assistance of SMOTE, the

most prominent predictors may be chosen with the assistance of the RFE, numerical predictors may be transformed with the assistance of Min-Max normalization. It will be able to ensure the balanced and clean data, as well as the optimization of deep learning pipeline.

g. *Flask-Based Deployment Module*

It is also a Web based module which is not very complex to access to a real time prediction system. Predictive rating engine can provide credit rating on real time basis, by using financial information uploaded and analysed by prediction engine. This module is scalable, does not need many resources to implement and produce in the case of financial institutions.

h. *Output & Reporting Module*

Prediction and risk classification coupled with confidence score are given to the users of this module. It also provides pictorial summaries where analysts can download reports and logs in an attempt to determine the data on the credit risks.

2. ALGORITHMS:

a) *Hybrid LSTM + Attention*

The Hybrid LSTM with Self-Attention model is the most predictive model of credit scores because the model shows all measures of assessment on the graph. It is very accurate as it can distinguish great and bad credit profile. The model scores high F1-score, accuracy and recall scores indicate that the model balances the false positives and negatives, and at the same time identifies the most significant patterns of time. The attention mechanism is what makes the model enhance its performance since it enhances the contextual awareness due to the emphasis on key financial activities. This improved representation is more accurate in prediction than the conventional machine learning models and the simple Hybrid LSTM.

b) *Proposed Hybrid LSTM*

The hybrid LSTM model is even poorer than a traditional machine learning method as it still is poorer than the Hybrid LSTM + Attention model. The fact that it has high accuracy and F1-score prove its ability to extrapolate long-term correlations of sequential financial data. The model is also useful to manage the imbalance in classes and the high-risk clients with lower false classification as the model has a high accuracy and recall. In cases where the attention process has not been used to narrow down the context, LSTM layers can store temporal behavioural patterns, i. e. spending and repayment history. This gives sound performance in credit risk evaluation.

c) *Random Forest*

Random Forest algorithm is superior in all the graph measures when compared to the two Hybrid LSTM models. When time or sequential financial data is used, F1-score, accuracy and recall scores show equal capacities of

prediction, yet display limitations. The limitation of the behavioral risk indicators that can be defined by Joe using the random Forest is that it cannot indicate time-specific trends although better in structured financial factors. It is still operational to be used as a reference format to credit scoring systems in which tabular data are used.

d) *XGBoost*

XGBoost will be just as accurate, precise and recalling and F1-score as a Random Forest, but will be slightly inferior to deep learning models. Its gradient-boosting algorithm does not support sequential data or dynamic behavioral data, however, it can capture nonlinear associations in the financial sector like Random Forest. This implies that XGBoost is predictive but gradual in its effectiveness to generate temporal credit risk indications. It is a weak performer when compared to LSTM-based models that take sequence of transactions into account, in addition to credit patterns over time and a good classifier of traditional financial data.

V. EXPERIMENTAL RESULTS

As it has been demonstrated, the Extension Hybrid LSTM with Self-Attention model outperforms other algorithms in an experiment. In the results, the proposed change can be more precise (89.87%), than the basic Hybrid LSTM (88.54%), the Random forest (87.22%), and the XGBoost (86.65%). This illustrates the more the self-attention mechanism is added to the model, the more it focuses on the time patterns of the financial patterns. Moreover, the F1-score measure is 84.55 which implies the increased accuracy-recall ratio and, consequently, the increased utility of the model to a real-world credit scoring application.

The Extension Hybrid LSTM + Attention model that achieved a 85.75% accuracy and an 83.51% recall can be used to recognize high-risk and low-risk customers. On the other hand, the standard Hybrid LSTM will only remember badly (77.55) and categorize (86.52) situations that can bring harm to human beings because of the non-attention-based prioritization of features. Random Forest and XGBoost perform slightly less well in terms of decreased recall and F1-score because it does not determine the temporal correlations between the financial behavior. The performance graph indicates the power, performance and suitability of the proposed extension to the current credit risk prediction systems through having an improved performance compared to all the current models in all assessment parameters.

Accuracy: One of the measures of reliability of the tests is the capacity of the tests to distinguish between people, who are healthy and those that are unhealthy. Reliability of a test can be found by using negative as well as the positive sides of the test. Following math:

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Precision is a measure of positive instances or accuracy of classification. To confirm, utilize:

$$Precision = \frac{TP}{(TP + FP)}$$

Recall: Resume of a model is the capability of an instance of tracing all the cases of a machine learning categorization. The comfort of the model determining the instances of the classes is provided by the quantity of correct positively forecasted observations to the total sum of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

F1-Score: F1 values are high indicating that the model of ML is right. Recall and accuracy are also more effective in improving model accuracy. Statistic of precision is used to tell the accuracy of a model in forecasting a collection of data.

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$

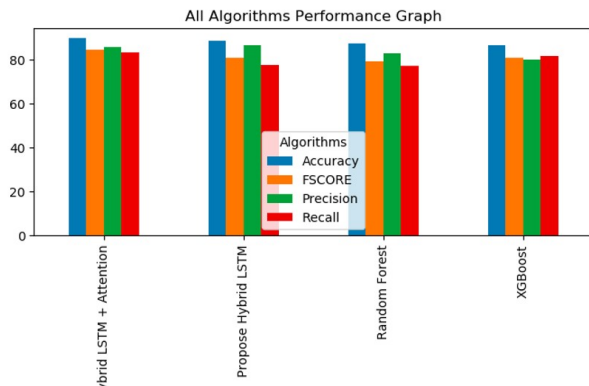


Fig 3 Performance Comparison graph

Algorithm Name	Accuracy	Precision	Recall	F1-Score
Random Forest	87.22	82.91	77.07	79.39
XGBoost	86.65	80.05	81.56	80.76
Propose Hybrid LSTM	88.54	86.52	77.55	80.82
Extension Hybrid LSTM + Attention	89.87	85.75	83.51	84.55

Table 1 Performance Comparison table

VI. CONCLUSION

Extension Hybrid LSTM model with Self-Attention is more precise in credit scores prediction as compared to the standard machine learning model and the simple architecture. The attention mechanism is most accurate with improved precision of 89.87 with a maximum accuracy of 89.87%, recall, and F1-score, which identifies important financial behavior and predicts time relationship. The findings indicate that, the model being considered has a greater capability to observe risk patterns than its traditional counterparts in terms of integration of attention. Data quality and balanced learning is enhanced by complex preparation approaches, including SMOTE, RFE and PCA. The proposed system is optimistic in the environment of the modern financial applications because of its robustness, scalability, and applicability in carrying out actual time credit risk evaluation.

REFERENCES

- [1] T. Schick and H. Schütze, “Exploiting cloze-questions for few-shot text classification and natural language inference,” in Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics, Main Volume, 2021. [Online]. Available: <https://arxiv.org/abs/2001.07676>
- [2] S. Hu, N. Ding, H. Wang, Z. Liu, J. Wang, J. Li, W. Wu, and M. Sun, “Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification,” in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2022. [Online]. Available: <https://arxiv.org/abs/2108.02035>
- [3] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, “Text classification via large language models,” in Proc. Findings Assoc. Comput. Linguistics (EMNLP), 2023. [Online]. Available: <https://arxiv.org/abs/2305.08377>
- [4] X. Han, W. Zhao, N. Ding, Z. Liu, and M. Sun, “PTR: Prompt tuning with rules for text classification,” AI Open, vol. 3, pp. 182–192, Jan. 2022.
- [5] Z. Wen and Y. Fang, “Augmenting low-resource text classification with graph-grounded pre-training and prompting,” in Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2023, pp. 506–516.
- [6] S. Garg and G. Ramakrishnan, “BAE: BERT-based adversarial examples for text classification,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020. [Online]. Available: <https://arxiv.org/abs/2004.01970>
- [7] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad-Khasmakhi, M. Chenaghlu, and J. Gao, “Deep learning-based text classification,” ACM Comput. Surveys, vol. 54, no. 3, pp. 1–40, 2021.
- [8] J. Chen, Z. Yang, and D. Yang, “MixText: Linguistically-informed interpolation of hidden space for semi-supervised text classification,” in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, 2020. [Online]. Available: <https://arxiv.org/abs/2004.12239>
- [9] Z. Jiang, M. Y. R. Yang, M. Tsirlin, R. Tang, Y. Dai, and J. J. Lin, “‘Lowresource’ text classification: A parameter-free classification method with compressors,” in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2023. [Online]. Available: https://aclanthology.org/2023.findingsacl.426/?trk=public_post_comment-text
- [10] S. Min, M. Lewis, H. Hajishirzi, and L. Zettlemoyer, “Noisy channel language model prompting for few-shot text classification,” in Proc. 60th Annu. Meeting Assoc.

- Comput. Linguistics (Long Papers), vol. 1, 2022. [Online]. Available: <https://arxiv.org/abs/2108.04106>
- [11] A. Karimi, L. Rossi, and A. Prati, "AEDA: An easier data augmentation technique for text classification," in Proc. Findings Assoc. Comput. Linguistics: EMNLP, 2021. [Online]. Available: <https://arxiv.org/abs/2108.13230>
- [12] S. Molitorys, B. Pilot, and K. Smyrak, "Text classification," in Proc. Int. Conf. Inf. Technol., 2024. [Online]. Available: https://www.researchgate.net/profile/V-Tampakas/publication/234812606_Text_classification_a_recent_overview/links/0c96051ee1dfd2a9f8000000/Text-classification-a-recent-overview.pdf
- [13] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, and L. Wang, "Every document owns its structure: Inductive text classification via graph neural networks," in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, 2020. [Online]. Available: <https://arxiv.org/abs/2004.13826>
- [14] A. Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, "A survey on text classification algorithms: From text to predictions," *Information*, vol. 13, no. 2, p. 83, Feb. 2022.
- [15] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, "A finetuned BERT-based transfer learning approach for text classification," *J. Healthcare Eng.*, vol. 2022, pp. 1–17, Jan. 2022.
- [16] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Hum. Res.*, vol. 5, no. 1, Dec. 2020. [Online].
- [17] Z. Wang, P. Wang, L. Huang, X. Sun, and H. Wang, "Incorporating hierarchy into text encoder: A contrastive learning approach for hierarchical text classification," in Proc. 60th Annu. Meeting Assoc. Comput. Linguistics (Long Papers), vol. 1, 2022. [Online]. Available: <https://arxiv.org/abs/2203.03825>
- [18] V. Dogra, S. Verma, Kavita, P. Chatterjee, J. Shafi, J. Choi, and M. F. Ijaz, "A complete process of text classification system using state-of-the-art NLP models," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–26, Jun. 2022.
- [19] M. Bayer, M.-A. Kaufhold, and C. Reuter, "A survey on data augmentation for text classification," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–39, Jul. 2023.
- [20] J. Zhou, C. Ma, D. Long, G. Xu, N. Ding, H. Zhang, P. Xie, and G. Liu, "Hierarchy-aware global model for hierarchical text classification," in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, 2020. [Online]. Available: <https://aclanthology.org/2020.acl-main.104/>
- [21] S. Soni, S. S. Chouhan, and S. S. Rathore, "TextConvoNet: A convolutional neural network based architecture for text classification," *Appl. Intell.*, vol. 53, no. 11, pp. 14249–14268, Jun. 2023.
- [22] M. Umer, Z. Imtiaz, M. Ahmad, M. Nappi, C. Medaglia, G. S. Choi, and A. Mehmood, "Impact of convolutional neural network and FastText embedding on text classification," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 5569–5585, Feb. 2023.
- [23] P. Atanasova, J. G. Simonsen, C. Lioma, and I. Augenstein, "A diagnostic study of explainability techniques for text classification," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-51518-7_7
- [24] K. Ding, J. Wang, J. Li, D. Li, and H. Liu, "Be more with less: Hypergraph attention networks for inductive text classification," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020. [Online]. Available: <https://arxiv.org/abs/2011.00387>
- [25] J. Chen, R. Zhang, Y. Mao, and J. Xu, "ContrastNet: A contrastive learning framework for few-shot text classification," in Proc. AAAI Conf. Artif. Intell., vol. 36, 2022, pp. 10492–10500.
- [26] A. S. Alammary, "BERT models for Arabic text classification: A systematic review," *Appl. Sci.*, vol. 12, no. 11, p. 5720, Jun. 2022.
- [27] X. Chen, P. Cong, and S. Lv, "A long-text classification method of Chinese news based on BERT and CNN," *IEEE Access*, vol. 10, pp. 34046–34057, 2022.
- [28] X. Liu, X. You, X. Zhang, J. Wu, and P. Lv, "Tensor graph convolutional networks for text classification," in Proc. AAAI Conf. Artif. Intell., vol. 34, 2020, pp. 8409–8416.
- [29] Y. Lin, Y. Meng, X. Sun, Q. Han, K. Kuang, J. Li, and F. Wu, "BertGCN: Transductive text classification by combining GNN and BERT," in Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP), 2021. [Online]. Available: <https://arxiv.org/abs/2105.05727>
- [30] B. Jang, M. Kim, G. Harerimana, S.-U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, p. 5841, Aug. 2020.