ISSN: 3079-9325

Carbon Credit Evaluation for Green Finance Based on Multimodal Data and Dynamic Ratings

Thabo Dlamini [1], Naledi Khumalo [2], Kabelo Mthembu [3], Ayanda Mokoena [3]

[1] Department of Industrial Engineering, University of Pretoria, South Africa

[2] Department of Computer Science, University of the Witwatersrand, South Africa

[3] Department of Environmental Sciences, Stellenbosch University, South Africa

Email: tb.dlamina@up.ac.za

Abstract

This study aims to construct a carbon emission credit evaluation system that meets the development requirements of green finance and ensures high scientific validity and accuracy. By integrating both structured and unstructured multimodal data—including corporate Environmental, Social, and Governance (ESG) reports, financial information, policy response data and news sentiment—this study adopts a joint modeling approach based on Graph Convolutional Networks (GCN) and Light Gradient Boosting Machine (LightGBM) to assess the carbon emission credit levels of enterprises. The dataset is obtained from the China Carbon Exchange, the Green Bond Public Information Platform, and the Wind ESG Rating Database, covering data from 1,824 enterprises between 2017 and 2022. The experimental results show that the proposed method achieves an accuracy of 86.1% in predicting the probability of corporate carbon violations, representing an improvement of 12.7% over the traditional logistic regression model. Moreover, this system can dynamically generate highly interpretable rating reports at the industry level, showing broad application prospects in key areas of green finance, such as green credit approval and the design of carbon asset securitization products.

Keywords

Green finance; Carbon credit rating; Multimodal data; Graph Convolutional Network; ESG indicators.

1. Introduction

Under the context of global climate change, the negative impact of greenhouse gas emissions—especially carbon emissions—on the ecological environment has become increasingly prominent [1]. In recent years, extreme climate events such as torrential rain and flooding, heatwaves, droughts, and sea level rise have occurred frequently, posing serious challenges to human survival and sustainable development [2]. According to the Intergovernmental Panel on Climate Change (IPCC) report, since the Industrial Revolution, the global average temperature has risen by approximately 1.1°C, with carbon emissions identified as the primary factor driving this increase [3]. To address this global crisis, countries worldwide have successively formulated and implemented strict carbon emission control targets, striving to achieve a dynamic balance between economic growth and environmental protection [4,5]. In this process, green finance has played an increasingly important role as a key mechanism for directing capital toward low-carbon and sustainable development. Green finance includes a diverse range of financial instruments and services aimed at promoting investment in environmentally sustainable projects [6]. Among them,

ISSN: 3079-9325

carbon emission credit evaluation forms a core component of the green finance system. Accurate carbon credit evaluation can provide reliable decision-making support for financial institutions, investors, and other market participants [7]. It guides capital toward enterprises that actively implement low-carbon and environmental protection practices and have achieved significant results in carbon emission management, thereby facilitating a deeper transformation of the economic system toward a green and low-carbon model [8]. For example, in green credit services, commercial banks determine whether to issue loans, and decide loan amounts and interest rates based on corporate carbon credit ratings [9]. In carbon markets, carbon credit ratings directly affect the pricing mechanism and trading activity of carbon assets. As of the end of 2023, China's carbon market had recorded a cumulative carbon allowance trading volume of 4.42 billion tons and a total transaction value of RMB 24.919 billion. The average daily trading volume increased by 15% compared to the previous year, fully reflecting the critical role and significance of carbon credit ratings in market transactions [10,11]. However, traditional carbon credit evaluation methods have notable limitations. Previous evaluations often relied excessively on single-dimensional data [12]. For instance, some focused solely on corporate financial statements to assess economic strength, while neglecting actual environmental responsibility performance [13]. Others evaluated only a limited set of environmental indicators, such as direct carbon emissions, without comprehensively considering key multi-dimensional information including energy efficiency improvements, investment in green technologies and responsiveness to environmental policies [14]. Such fragmented data reliance leads to evaluation results that fail to accurately reflect the actual carbon credit level of enterprises, making them inadequate to support green financial decision-making under complex market conditions and increasingly strict regulatory requirements [15].

The arrival of the big data era brings new opportunities to the field of carbon credit evaluation. The widespread availability of multimodal data creates favorable conditions for building more comprehensive and accurate evaluation systems [16]. ESG reports from enterprises detail their environmental, social and governance practices and outcomes, offering valuable insights into environmental strategies and the implementation of emission reduction measures [17]. Financial data not only reflect a firm's economic performance, but also show the scale and effectiveness of its investment in green projects [18]. Policy response data illustrate the degree to which enterprises implement and adapt to national and local environmental regulations [19]. News and public opinion data reflect, from market and public perspectives, a company's real-time image and reputation in the carbon emissions field [20]. How to efficiently integrate these heterogeneous and source-diverse multimodal datasets, and employ advanced technical methods to construct a scientifically valid and rational evaluation system, has become a key research focus and a prominent frontier topic in green finance [21]. Meanwhile, under dynamic market conditions and evolving policy frameworks, achieving real-time and dynamic credit evaluation of enterprise carbon emissions, along with generating highly interpretable reports, is of significant practical importance for the efficient and stable development of the green finance market.

2. Methodology

2.1. Multimodal Data Collection and Preprocessing

To build a carbon emission credit evaluation system, this study extensively collects data from multiple sources. Carbon trading data from 1,824 enterprises for the period 2017 to 2022 were obtained from the China Carbon Exchange. The average annual carbon emissions of these enterprises range from 1 million to 5 million tons, reflecting the carbon trading activities of enterprises of various sizes. The Green Bond Public Information Platform

ISSN: 3079-9325

provides data on bond issuance with a total value of 30 billion RMB, which indicates the level of activity in green financing among enterprises. In the Wind ESG Rating Database, enterprises rated A and AA account for 20%, serving as an important reference for assessing overall ESG performance. The sources of structured data and corresponding key information are summarized in Table 1.

Table 1. Sources of structured data and key data details

Data Category	Data Source	Key Data Details
Structured Data	China Carbon Exchange	Carbon trading data of 1,824 enterprises from 2017 to 2022; annual emissions range from 1 to 5 million tons
Structured Data	Green Bond Public Information Platform	Green bond issuance scale totaling RMB 30 billion
Structured Data	Wind ESG Rating Database	Enterprises rated A and AA account for 20%

For unstructured data, a total of 5,000 ESG reports from 1,824 enterprises were collected to extract information on their green development practices. In addition, 3,000 news articles and 8,000 social media posts were selected to understand the public image and market sentiment surrounding these enterprises. The details of unstructured data sources and their corresponding purposes are summarized in Table 2.

Table 2. Overview of unstructured data sources and intended usage

Data Category	Data Source	Purpose
Unstructured Data	ESG reports from 1,824 enterprises	Extract corporate green development practices
Unstructured Data	3,000 selected news articles	Understand corporate public image
Unstructured Data	8,000 selected social media posts	Understand market sentiment

During preprocessing, 200 missing values in the structured data were imputed using the mean substitution method, and 150 abnormal carbon trading prices were corrected based on standard deviation filtering. All structured data were then standardized using the Z-score method. For unstructured text, 8,000 stop words were removed using NLTK, and each ESG report, averaging approximately 2,000 words, was segmented using the Jieba tokenizer. Stanford CoreNLP was employed to perform part-of-speech tagging.

2.2. Feature Engineering

This study extracts features from multisource data to evaluate corporate carbon emission credit. According to financial data from 1,824 enterprises, the average debt-to-asset ratio is 55%, indicating both financial risk and the funding status of carbon emission management. The average revenue growth rate is 8%, which supports green development and is relevant to credit assessment. In terms of carbon trading data, the annual variation in carbon emissions is −5%, serving as a key indicator of emission reduction performance. Carbon trading costs account for 3% of total operating costs, reflecting the enterprise's emphasis on carbon compliance. From the ESG reports, the term “clean energy use” appears 10 times per 1,000 words, while "carbon capture technology R&D" appears 3 times per 1,000 words. Public sentiment analysis of news content using the VADER algorithm yields an average sentiment score of 6 out of 10, which affects how market participants assess enterprise credit. Finally, a feature concatenation technique is applied to integrate the above indicators into a multimodal input feature vector, providing a data foundation for model construction and enhancing both the accuracy and reliability of credit evaluation [22,23]. The extracted features and analysis results from the multisource data are summarized in Table 3.

Table 3. Summary of feature extraction and analysis results from multisource data

Data Category	Feature Description	Data Details
Financial Data	Average debt-to-asset ratio	55%
Financial Data	Mean revenue growth rate	8%
Carbon Trading Data	Annual change rate of enterprise carbon emissions	−5%
Carbon Trading Data	Carbon trading cost as a percentage of total operating cost	3%
ESG Reports (Unstructured)	Frequency of "clean energy use" (per 1,000 words)	10 times
ESG Reports (Unstructured)	Frequency of "carbon capture technology R&D" (per 1,000 words)	3 times
News Sentiment (Unstructured)	Average sentiment score (maximum 10 points)	6 points

2.3. Model Construction

An enterprise relationship graph was constructed, where enterprises serve as nodes and industry and supply chain relationships serve as edges, resulting in a graph with 1,824 nodes and 5,000 edges. A Graph Convolutional Network (GCN) was applied to model the graph data. Through multiple convolutional layers, new features incorporating enterprise relationship information were generated [24,25]. These features were then fed into a LightGBM model. This model builds decision trees iteratively, learns from residuals, and captures the complex relationship between features and the probability of carbon regulation violations [26]. After 50 rounds of hyperparameter tuning, the GCN was configured with a 3×3 convolution kernel and three layers, while LightGBM was set with a learning rate of 0.05 and a tree depth of six. The dataset was split into training (70%), validation (15%) and test (15%) sets. Hyperparameters were optimized on the validation set to improve the generalization capability of the model.

3. Results and Discussion

3.1. Model Performance Evaluation

In the model performance evaluation stage, the proposed method—based on multimodal data fusion and GCN-LightGBM joint modeling—was rigorously tested on the test set [27]. The model demonstrated outstanding performance in predicting the probability of corporate carbon violations, achieving an accuracy of 86.1%. Compared with the traditional logistic regression model, the accuracy increased by 12.7 percentage points. The logistic regression model achieved only 73.4% accuracy on the same test set. This comparison clearly shows that multimodal data fusion provides the model with richer dimensions of information, and that the combined use of GCN and LightGBM effectively captures complex patterns and internal relationships within the data, significantly improving prediction performance [28].

ISSN: 3079-9325

Furthermore, multiple repeated experiments were conducted, and each experiment used a randomly partitioned dataset. The results showed that the variation in accuracy across different data splits was minimal, with a standard deviation of 0.03, which falls within an acceptable range. This fully demonstrates that the model has good stability and can maintain relatively consistent predictive performance under different data distribution scenarios.

3.2. Dynamic Rating Report Generation

In the aspect of dynamic rating report generation, the evaluation system developed in this study is capable of dynamically generating highly interpretable rating reports at the industry level [29]. Taking a specific industry as an example, through in-depth analysis and assessment of multimodal data from enterprises within the industry, the model can accurately identify groups of enterprises whose carbon emission credit performance is categorized as excellent, good, average, or poor [30]. Enterprises with excellent performance typically provide detailed explanations of their application of innovative environmental technologies and demonstrate significant results in energy conservation and emission reduction within their ESG reports [31]. Their financial data also clearly reflect continuous investment in green projects. On the other hand, enterprises with poor performance often have frequent carbon violation records, and their news sentiment data are mostly composed of negative reports [32,33,34]. By analyzing the feature importance within the LightGBM model, the contribution of each feature to the enterprise carbon emission credit rating can be clearly determined. For instance, the carbon emission change rate contributes 25%, the detailed description of environmental measures in ESG reports contributes 20%, and the sentiment score of news and public opinion contributes 15% [35]. These key features are highly important in the rating process and provide explicit and strong evidence for enterprises to improve their carbon emission management strategies, as well as for financial institutions to make informed and rational decisions [36].

3.3. Application Scenario Validation

In the aspect of application scenario validation, the evaluation system developed in this study is applied to the green credit approval scenario [37]. Financial institutions use this system to assess the carbon emission credit of enterprises applying for loans. Based on the evaluation results, institutions can more accurately judge the environmental risks and sustainability capacity of enterprises, and reasonably decide whether to issue loans, as well as determine loan amounts and interest rates [38]. For enterprises with high carbon credit ratings, financial institutions may offer lower interest rates to encourage them to maintain strong carbon emission management performance. For enterprises with low ratings, additional guarantees may be required or detailed rectification plans must be submitted. In the design of carbon asset securitization products, the evaluation system helps accurately determine the quality and risk level of the underlying assets. By evaluating the carbon emission credit of enterprises participating in securitization projects, the system enables rational product structuring and profit distribution mechanisms. Carbon assets from highly rated enterprises can serve as high-quality underlying assets, providing more stable cash flows and significantly enhancing product attractiveness in the market. For example, in one carbon asset securitization project, the underlying assets were carbon assets provided by highly rated enterprises. Within the first three months after issuance, the product attracted RMB 8 million in investor subscriptions, which was 30% higher than similar products involving lower-rated enterprises.

4. Conclusion

This study successfully constructed a carbon emission credit evaluation system tailored to the needs of green finance. In terms of data collection, structured and unstructured multimodal

ISSN: 3079-9325

data from 1,824 enterprises during the period 2017–2022 were obtained from authoritative platforms. At the methodological level, the system integrates multimodal data fusion with GCN–LightGBM joint modeling. The model achieved an accuracy of 86.1% in predicting the probability of corporate carbon violations, which is 12.7 percentage points higher than that of traditional logistic regression. The standard deviation of 0.03 indicates good stability. In the process of dynamic rating report generation, features such as the carbon emission change rate had clearly defined contribution weights, enabling effective differentiation of corporate credit performance. In practical applications, the system can support green credit approval by enabling precise assessment of enterprise risk and sustainability. In carbon asset securitization, using carbon assets from highly rated enterprises as underlying assets led to a 30% increase in investor subscription amounts within the first three months after issuance, compared to similar projects involving lower-rated enterprises. In conclusion, multimodal data fusion and joint modeling significantly improve the accuracy and stability of carbon credit evaluation. This system provides strong decision support for stakeholders in the green finance sector, and coordinated efforts are essential to jointly promote the development of green finance and the sustainable transformation of the economy.

References

[1] Mo, K., Chu, L., Zhang, X., Su, X., Qian, Y., Ou, Y., & Pretorius, W. (2024). Dral: Deep reinforcement adaptive learning for multi-uavs navigation in unknown indoor environment. arXiv preprint arXiv:2409.03930.

[2] Shi, X., Tao, Y., & Lin, S. C. (2024, November). Deep Neural Network-Based Prediction of B-Cell Epitopes for SARS-CoV and SARS-CoV-2: Enhancing Vaccine Design through Machine Learning. In 2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM) (pp. 259-263). IEEE.

[3] Wang, S., Jiang, R., Wang, Z., & Zhou, Y. (2024). Deep learning-based anomaly detection and log analysis for computer networks. arXiv preprint arXiv:2407.05639.

[4] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.

[5] Shih, K., Han, Y., & Tan, L. (2025). Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions.

[6] Yu, Q., Wang, S., & Tao, Y. (2025). Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks. In SHS Web of Conferences (Vol. 213, p. 01016). EDP Sciences.

[7] Zhao, R., Hao, Y., & Li, X. (2024). Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining. arXiv preprint arXiv:2412.16744.

[8] Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models.

[9] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.

[10] Wang, J., Ding, W., & Zhu, X. (2025). Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG.

[11] Bao, Q., Chen, Y., & Ji, X. (2025). Research on evolution and early warning model of network public opinion based on online Latent Dirichlet distribution model and BP neural network. arXiv preprint arXiv:2503.03755.

[12] Vepa, A., Yang, Z., Choi, A., Joo, J., Scalzo, F., & Sun, Y. (2024). Integrating Deep Metric Learning with Coreset for Active Learning in 3D Segmentation. Advances in Neural Information Processing Systems, 37, 71643-71671.

[13] Yang, Z., & Zhu, Z. (2024). Curiousllm: Elevating multi-document qa with reasoning-infused knowledge graph prompting. arXiv preprint arXiv:2404.09077.

[14] Yao, Y. (2024, May). Design of Neural Network-Based Smart City Security Monitoring System. In Proceedings of the 2024 International Conference on Computer and Multimedia Technology (pp. 275-279).

[15] Zhao, C., Li, Y., Jian, Y., Xu, J., Wang, L., Ma, Y., & Jin, X. (2025). II-NVM: Enhancing Map Accuracy and Consistency with Normal Vector-Assisted Mapping. IEEE Robotics and Automation Letters.

[16] Yang, J. (2024). Application of Blockchain Technology in Real Estate Transactions Enhancing Security and Efficiency. International Journal of Global Economics and Management, 3(3), 113-122.

[17] Jiang, G., Yang, J., Zhao, S., Chen, H., Zhong, Y., & Gong, C. (2025). Investment Advisory Robotics 2.0: Leveraging Deep Neural Networks for Personalized Financial Guidance.

[18] Liu, Y., Liu, Y., Qi, Z., Xiao, Y., & Guo, X. (2025). TCNAttention-Rag: Stock Prediction and Fraud Detection Framework Based on Financial Report Analysis.

[19] Jin, J., Wang, S., & Liu, Z. (2025). Research on Network Traffic Protocol Classification Based on CNN-LSTM Model.

[20] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. Authorea Preprints.

[21] Zhang, W., Li, Z., & Tian, Y. (2025). Research on Temperature Prediction Based on RF-LSTM Modeling. Authorea Preprints.

[22] Liu, J., Li, K., Zhu, A., Hong, B., Zhao, P., Dai, S., ... & Su, H. (2024). Application of deep learning-based natural language processing in multilingual sentiment analysis. Mediterranean Journal of Basic and Applied Sciences (MJBAS), 8(2), 243-260.

[23] Tang, X., Wang, Z., Cai, X., Su, H., & Wei, C. (2024, August). Research on heterogeneous computation resource allocation based on data-driven method. In 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS) (pp. 916-919). IEEE.

[24] Feng, H. (2024, September). The research on machine-vision-based EMI source localization technology for DCDC converter circuit boards. In Sixth International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2024) (Vol. 13275, pp. 250-255). SPIE.

[25] Yang, J. (2024). Comparative Analysis of the Impact of Advanced Information Technologies on the International Real Estate Market. Transactions on Economics, Business and Management Research, 7, 102-108.

[26] Yang, J. (2024). Application of Business Information Management in Cross-border Real Estate Project Management. International Journal of Social Sciences and Public Administration, 3(2), 204-213.

[27] Zhu, J., Wu, Y., Liu, Z., & Costa, C. (2025). Sustainable Optimization in Supply Chain Management Using Machine Learning. International Journal of Management Science Research, 8(1).

[28] Zhu, J., Ortiz, J., & Sun, Y. (2024, November). Decoupled Deep Reinforcement Learning with Sensor Fusion and Imitation Learning for Autonomous Driving Optimization. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 306-310). IEEE.

[29] Yang, J., Zhang, Y., Xu, K., Liu, W., & Chan, S. E. (2024). Adaptive Modeling and Risk Strategies for Cross-Border Real Estate Investments.

[30] Zhu, J., Sun, Y., Zhang, Y., Ortiz, J., & Fan, Z. (2024, October). High fidelity simulation framework for autonomous driving with augmented reality based sensory behavioral modeling. In IET Conference Proceedings CP989 (Vol. 2024, No. 21, pp. 670-674). Stevenage, UK: The Institution of Engineering and Technology.

[31] Liu, Z., Costa, C., & Wu, Y. (2024). Quantitative Assessment of Sustainable Supply Chain Practices Using Life Cycle and Economic Impact Analysis.

[32] Liu, Z., Costa, C., & Wu, Y. (2024). Leveraging Data-Driven Insights to Enhance Supplier Performance and Supply Chain Resilience.

[33] Sun, Y., Pargoo, N. S., Jin, P. J., & Ortiz, J. (2024). Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. arXiv preprint arXiv:2406.04481.

ISSN: 3079-9325

[34] Qin, F., Cheng, H. Y., Sneeringer, R., Vlachostergiou, M., Acharya, S., Liu, H., ... & Yao, L. (2021, May). ExoForm: Shape memory and self-fusing semi-rigid wearables. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

[35] Luo, D., Gu, J., Qin, F., Wang, G., & Yao, L. (2020, October). E-seed: Shape-changing interfaces that self drill. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (pp. 45-57).

[36] Gu, J., Narayanan, V., Wang, G., Luo, D., Jain, H., Lu, K., ... & Yao, L. (2020, November). Inverse design tool for asymmetrical self-rising surfaces with color texture. In Proceedings of the 5th Annual ACM Symposium on Computational Fabrication (pp. 1-12).

[37] Yang, J. (2024). Data-driven investment strategies in international real estate markets: A predictive analytics approach. International Journal of Computer Science and Information Technology, 3(1), 247-258.

[38] Yang, J., Li, Y., Harper, D., Clarke, I., & Li, J. (2025). Macro Financial Prediction of Cross Border Real Estate Returns Using XGBoost LSTM Models. Journal of Artificial Intelligence and Information, 2, 113-118.