AI PATENT IDENTIFICATION IN A-SHARE MANUFACTURING ENTERPRISES BASED ON CNN-TRANSFORMER

Haoyue CHEN[1]*, Qijian PAN[2], Jiayue ZHANG[3]

Accurate identification and classification of artificial intelligence (AI) patents necessitate advanced computational methods capable of processing complex textual data. In order to improve the accuracy and effectiveness of AI patent analysis, this study presents a revolutionary deep learning model that combines Transformer topologies with Convolutional Neural Networks (CNN). Using a large dataset of AI patents published by A-share listed manufacturing companies between 2014 and 2023, we compare the CNN-Transformer model to well-known deep learning models like standard CNNs and Long Short-Term Memory networks (LSTM) as well as conventional machine learning algorithms like Decision Trees, Random Forests, and Support Vector Machines (SVM). Experimental results reveal that the CNN-Transformer model achieves superior performance, attaining an accuracy of 90.75%, precision of 91.10%, recall of 93.20%, F1 score of 90.97%, and AUC of 96.56%, thereby significantly outperforming all comparative models. These findings demonstrate the model's exceptional capability in handling complex patent text analysis, highlighting its potential as a powerful tool for AI patent classification tasks.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Convolutional Neural Networks, Transformer

1. Introduction

The incorporation of AI technology has had a significant influence on a number of sectors in the age of fast AI advancement, resulting in an unparalleled surge in technical development and innovation [1]. This surge in AI-driven innovation has resulted in a substantial increase in complex textual data, particularly in patent documents that encapsulate cutting-edge technological advancements [2]. Efficiently processing and analyzing this large-scale, unstructured textual data presents significant challenges for traditional computational methods.

Accurate classification of patent texts is essential for understanding

[1] School of Economics, Guangdong University of Technology, Guangzhou, China, e-mail: 18024811060@163.com
[2] School of Economics, Guangdong University of Technology, Guangzhou, China, e-mail: 15521076878@163.com
[3] School of Economics, Guangdong University of Technology, Guangzhou, China, e-mail: 1653766283@qq.com

technological trends, informing policy decisions, and guiding strategic business initiatives [3]. However, the intricate language structures, specialized terminologies, and sheer volume of patent documents make them difficult to analyze using conventional machine learning algorithms. Techniques like SVM, Decision Trees, and Random Forests have demonstrated limitations: SVM has too high dimension, high calculation cost, slow operation. [4,5]; Decision Trees are sensitive to training data and prone to overfitting [6]; Random Forests suffer from weaker model interpretability and difficulty in handling imbalanced datasets [7].

Tools like CNN and LSTM from deep learning are used to resolve certain challenges [8]. While CNNs are good at identifying local features, the use of pooling layers might cause contextual data to be discarded, and they often fail to consider the interaction between local and global aspects [9, 10]. LSTMs are capable of modeling sequential data but are prone to overfitting and involve high computational costs [11]. Although BERT excels at capturing long-distance dependencies, its capacity to model locally fine-grained patterns is relatively limited. In contrast, BERT-CNN has an excessive number of parameters, which leads to a prolonged training duration. These limitations underscore the need for more sophisticated computational models that can effectively capture both local and global semantic relationships within complex textual data.

This paper proposes an inventive CNN-Transformer approach to overcome these issues by fusing Transformer's global context modeling advantage with CNN's local feature identification capability. The Transformer component excels at capturing long-range dependencies and semantic nuances, which are critical for accurately classifying complex patent documents. By integrating these two architectures, the proposed model aims to overcome the deficiencies of existing methods and enhance text classification performance.

In addition to alternative deep learning models and conventional machine learning techniques, the study trains and assesses the suggested model using a manually annotated patent dataset based on the AI lexicon created by Yao et al. (2024) [12]. Model performance is tested using metrics like accuracy, recall, F1 score, and AUC. Experimental results indicate that the CNN-Transformer model significantly outperforms existing methods in classifying patent texts, demonstrating its effectiveness in handling complex textual analysis tasks.

Furthermore, to validate the practical applicability of the proposed model, it is employed to classify patents from A-share listed manufacturing enterprises between 2014 and 2023, identifying AI-related patents among them. In addition to demonstrating the model's resilience in practical situations, this application offers insightful information about how widely AI technology is being used in these businesses.

The following structure is used in this paper: Section 2 recalling existing computational methods for text classification and their limitations; Section 3 presents the research design, including data acquisition, preprocessing, and model construction; Section 4 provides empirical analysis and results, highlighting the performance of the proposed model; and Section 5 summarizes the study and examines avenues for future exploration.

2. Literature Review

2.1 Machine Learning

Machine learning, one of the core directions of AI, has been broadly employed in agriculture, logistics, and communication. In Natural Language Processing (NLP), machine learning technologies have been more broadly utilized to handle challenging tasks like text classification, abstraction, and emotion analysis [13]. Its significant advantages in processing unstructured data have led to extensive use in enterprise risk prediction, performance evaluation, and decision support [14]. These empirical studies offer an excellent foundation for the approaches chosen in this study, emphasizing machine learning's capacity to gauge the extent of enterprise-level adoption of AI technology.

SVM, a traditional machine learning technique founded on statistical learning theory, works especially well with high-dimensional, small-sample, and nonlinear data. They have been broadly utilized in areas such as image recognition, fault diagnosis, and text classification [5]. However, SVM exhibits considerable computational complexity, high computational cost and slow operation when dealing with large-scale data. In addition, multi-class problems require manual selection of kernel functions, which increases the complexity of model construction [4].

Decision Trees are widely used due to their low preprocessing requirements and ability to handle mixed-type data. However, Decision Trees react strongly to slight fluctuations in the training data, causing them to be prone to overfitting [6]. To overcome this limitation, Random Forests enhance model stability and accuracy by integrating multiple Decision Trees and employing voting or averaging mechanisms for output, thereby reducing the risk of overfitting [7].

CNN are extensively applied in the field of NLP for text analysis tasks. A CNN-based approach to text classification was presented by Guo et al. (2019). It uses Word2vec technology to convert text into numerical vectors and use convolutional kernels of different sizes to extract features. These features are then processed through pooling and classification layers to achieve efficient classification. The model captures local text features and optimizes computational load, thus enhancing training efficiency [9]. Even so, pooling layers may overlook the interplay between local and global information, leading to a partial loss of contextual information [10].

To address the limitations of CNNs in handling the contextual associations in long text sequences and the gradient reduction problem encountered during training by classic Recurrent Neural Networks face during training while processing lengthy sequences, Hochreiter and Schmidhuber (1997) invented LSTM. LSTMs can efficiently capture long-range dependencies, significantly alleviating the issues of gradient vanishing and explosion during training [11]. However, this model is prone to overfitting and incurs high computational costs.

BERT uses a Transformer pre-trained bidirectional encoder to improve language understanding through context-aware feature extraction and a deep bidirectional architecture [15]. Transformer is an important structure of PFM in the fields of NLP, CV and GL. For NLP, Transformer can help solve the long-distance dependency problem when processing sequential input data [16]. ChineseBERT, designed for Chinese characteristics, further expands the BERT architecture and enhances the ability to handle Chinese semantic ambiguity and homophones by integrating multimodal information of Chinese character shapes and pinyin [17]. Although BERT is good at capturing long-range dependencies, it has limited ability to model local fine-grained patterns [18].

BERT-CNN achieves high classification accuracy in short text sentiment analysis scenarios by combining BERT's deep semantic representation capabilities with CNN's local feature extraction mechanism [19]. However, this model is limited by the local receptive field characteristics of convolutional neural networks and is unable to effectively capture long-distance semantic dependencies [20].

Given the aforementioned limitations of existing models in processing complex text data, this study proposes an innovative deep learning model, CNN-Transformer. This model aims to fully utilize the strengths of CNNs in extracting local characteristics and the capabilities of Transformers in capturing overall dependencies and long-distance information. By integrating these two architectures, the CNN-Transformer model significantly enhances classification accuracy and efficiency in complex patent text classification tasks, overcoming the shortcomings of single models in feature extraction and context understanding.

Table 1

Literature Review of Machine Learning Algorithms

Machine Learning Algorithm	Advantages	Disadvantages	References
SVM	Suitable for nonlinear, small-sample, and high-dimensional data; widely used in image recognition, fault diagnosis, and text classification	high computational complexity for large-scale data; requires manual selection of kernel functions	Jiang-hua, 2002; Li et al., 2022
Decision Tree	Low preprocessing requirements; can handle mixed-type data	Sensitive to training data; prone to overfitting	Amro et al., 2021
Random Forest	Enhances model stability and accuracy; reduces risk of	Weaker model interpretability; difficult to	Sun et al., 2024

	overfitting	handle imbalanced datasets	
CNN	Effectively captures local text features; reduces computational complexity; accelerates training	Pooling layers may lead to loss of contextual information; neglects interaction between local and global information	Guo et al., 2019; Ma et al., 2023
LSTM	Effectively learns long-distance dependencies; alleviates gradient vanishing and explosion problems	Prone to overfitting; high computational costs	Hochreiter, 1997
BERT	Good at capturing long-distance dependencies and can simultaneously capture bidirectional contextual information of text	Limited ability to model local fine-grained patterns	Jacob et al,2019; Sun et al,2021; Chen et al,2022
BERT-CNN	Fine-grained sentiment classification has significant accuracy and is suitable for sentiment analysis of short texts	The model contains an excessive number of parameters, resulting in excessively long training durations.	Abas et al,2022; Rong et al,2023

The primary machine learning algorithms covered in this study's literature review are compiled in table 1 above, which also includes a list of pertinent references and highlights each algorithm's benefits and drawbacks. Through comparative analysis, the limitations of existing models in handling complex text data become evident, providing theoretical support and directions for improvement for the proposed CNN-Transformer model in this research.

2.2 Methods for Measuring the Level of AI Technology Adoption

Existing studies employ various methods to determine the capability level of AI technology adoption in enterprises, with questionnaire surveys, keyword analysis, and proxy variable approaches being the primary techniques.

The questionnaire survey method is commonly used to assess AI technology application by collecting feedback from internal managers of enterprises. For example, Liu et al. (2022) constructed an intelligent apparel ecosystem under the context of digital transformation based on platform ecosystem theory using questionnaire surveys and proposed a customer trust model based on human-computer interaction technology [21]. Sharma et al. (2022) measured the level of AI technology adoption by surveying the attitudes and perspectives of machine learning (ML) and AI researchers [22]. However, the questionnaire survey method relies heavily on the subjective evaluations of enterprise managers and is susceptible to the "halo effect," where the order of questions and other factors may influence respondents' answers [23].

The keyword method has been used extensively as a useful instrument for gauging the degree of technological innovation in AI within businesses. For example, Liu et al. (2022) created an indicator to gauge the degree of AI development in businesses by counting the AI-related phrases they extracted from the financial documents of Chinese listed companies using web crawler technology [24]. Nguyen et al. (2022) utilized the International Federation of Intellectual Property (IFIClaims) patent dataset and applied the Panel Smooth Transition Regression (PSTR) model to analyze the relationship between AI and unemployment under different inflation levels [25]. Zhang and Peng (2024) generated an AI dictionary using machine learning methods and constructed an enterprise-level AI indicator through text analysis of annual data from listed companies [26]. Although the keyword method effectively identifies trends, its insufficient understanding of context may lead to biases in semantic and sentiment interpretation [13].

The proxy variable method indirectly measures the level of AI technology adoption by using substitute variables that are difficult to directly observe or quantify. In order to scientifically assess the energy efficiency of micro-enterprises and investigate the role of AI in it, Yang and Wang (2023) use panel data from 2013 to 2015 that integrates enterprise-level data from a thorough survey of over 110,000 manufacturing enterprises in Guangdong Province with the global count of industrial robots by industry from the International Federation of Robotics (IFR) [10]. The concentration of industrial robots at the provincial agricultural level was used by Lee et al. (2024) to gauge the application and development level of AI in agriculture [27]. Furthermore, Liu et al. (2022) determined the percentage of enterprise output value to the industry's overall output value in order to gauge the extent of AI adoption [28]. However, the proxy variable method depends on the completeness of databases and the accuracy of variable definitions and typically only identifies data at the national and industry levels, limiting in-depth analysis of digital technology characteristics.

In summary, although existing methods provide important insights into measuring the level of AI technology adoption in enterprises, each has certain limitations. The questionnaire survey method primarily captures the short-term technological adoption of enterprises. The keyword analysis method can effectively identify trends but struggles to fully understand the context, potentially leading to semantic and sentiment interpretation biases [15]. The proxy variable method is limited in its ability to comprehensively and deeply characterize AI technology. Therefore, this study proposes to use a deep learning model that combines CNN and Transformer architectures to efficiently identify and classify corporate AI technology patents through text analysis methods, improve the tracking of corporate strategic changes, and deepen the application of machine learning in text semantic understanding. Table 2 offers a methodical summary of the approaches used in previous studies, highlighting their benefits and drawbacks.

Table 2

Overview of Existing Research on Measuring AI Levels

Research Method	Data Source	Method	Advantages	Limitations	References
Questionnaire Survey	Questionnaire Surveys	Collect subjective evaluations of AI adoption from managers by designing and distributing questionnaires	Captures short-term technological adoption within enterprises	Easily influenced by the "halo effect," high subjectivity	Liu & Li, 2022; Sharma et al., 2022; Xie et al., 2024
Keyword Analysis	Annual Reports of Listed Companies, Patent Datasets	Extract AI-related keywords and construct AI adoption indicators through counting and statistical analysis	Effectively identifies technological development trends	Insufficient understanding of context may lead to semantic and sentiment biases	Liu et al., 2022; Nguyen & Vo, 2022; Zhang & Peng, 2024; Wankhade et al., 2022
ffigureProxy Variable Method	Industrial Robot Usage Data, Enterprise Output Data	Use substitute variables (e.g., robot stock, robot density) to indirectly measure AI adoption levels	Utilizes existing data for indirect measurement, suitable for cross-country or cross-industry analysis	Strong dependency on data completeness and variable definition accuracy, difficult to deeply characterize AI technology	Yang & Wang, 2023; Lee et al., 2024; Liu, Qian, Yang & Yang, 2022

3. Research Design

3.1 Sample Selection and Data Sources

This study chooses companies listed in the A-share market manufacturing enterprises over the past decade (2014-2023) as the initial research sample. The data for the sample companies are categorized into two types: firstly, basic company information is obtained from the CSMAR database; secondly, patent text information of listed companies is sourced from the WinGo Financial Text Data Platform, including the company name, patent application date, and the abstract, description, and claims of the patent documents. Additionally, to minimize the impact of financial anomalies, companies classified as ST and *ST are excluded from the initial sample. After sorting, this study obtained a non-equilibrium data set containing 3533 manufacturing enterprises, with a total of 986296 samples, of which 3000 positive samples and 7000 negative samples.

3.2 Text Vectorization and Representation

To effectively input textual data into machine learning and deep learning models, this study employs two different text vectorization methods. For traditional machine learning models, the TfidfVectorizer from the scikit-learn library is used for vectorization. By computing Term Frequency-Inverse Document Frequency (TF-IDF), TfidfVectorizer produces sparse high-dimensional feature vectors that accurately capture significant lexical information in text.

For deep learning models, a text embedding method is utilized. Specifically, each Chinese character is ranked by frequency, and characters with a frequency below 5% are removed, retaining a total of 4,906 high-frequency characters. Considering that over 95% of the texts (including patent titles and abstracts) are fewer than 256 characters in length, the texts are uniformly processed to a length of 256 characters, truncating any excess and padding shorter texts. This process ensures uniformity of input vectors and enhances the efficiency of model training.

3.3 Model Construction and Implementation

This study evaluates and compares the performance of traditional machine learning models—Decision Trees, Random Forests, and SVM—as well as deep learning models including LSTM, BERT, CNN-BERT, and a proposed CNN-Transformer hybrid model. Traditional models are implemented using the scikit-learn library, with hyperparameter tuning conducted through grid search.

Deep learning models use a 128-dimensional embedding layer to convert input character vectors into dense representations. The CNN baseline consists of two convolutional layers with 64 filters and a kernel size of 3, designed to capture local textual features. The LSTM model includes two stacked layers with 64 hidden units to capture long-range dependencies. Transformer-based models include BERT, which uses a pretrained TFBertModel to extract the hidden state of the CLS token, followed by a fully connected layer with 128 units and a sigmoid activation for binary classification. CNN-BERT enhances this by applying a one-dimensional convolution with 128 filters and a kernel size of 3 to the final hidden states, followed by global max pooling and a dense layer with 128 units.

All models are implemented and evaluated under consistent experimental settings to ensure a fair and rigorous comparison.

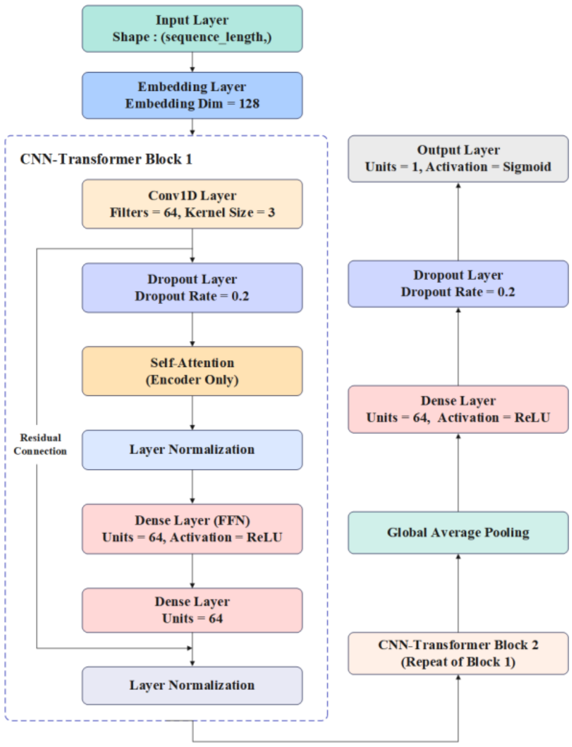

Fig. 1 CNN-Transformer Model Architecture

The proposed architecture consists of two repeatedly stacked modules, each comprising a convolutional layer followed by a Transformer encoder layer. In this design, only the encoder component of the Transformer is utilized, explicitly omitting the decoder structure, as the text classification task does not require sequence generation or reconstruction.

Each convolutional layer employs 64 filters with a kernel size of 3, effectively capturing local patterns in the input sequence. These local features are then processed by the Transformer encoder layer, which includes a multi-head self-attention mechanism with two attention heads and a 64-dimensional hidden state. This configuration enables the model to capture both local and global dependencies within the sequence, enhancing the contextual representation.

The input sequence is first converted into continuous embeddings, with positional encodings added to retain sequential information. These enriched embeddings pass through the stacked CNN-Transformer blocks, where the convolutional layers extract localized features, and the attention layers model long-range semantic relationships. Unlike sequence-to-sequence tasks such as machine translation that rely on the decoder to generate sequential outputs, text classification only requires a compact semantic representation of the entire input.

Therefore, the output of the final encoder layer is passed through a global average pooling layer to produce a fixed-length representation, which is then fed into a fully connected layer and activated using a sigmoid function to yield a binary classification result. This architecture, illustrated in Fig. 1, offers a simplified and computationally efficient model structure that is well-aligned with the objective of the classification task.

The hybrid structure makes full use of the advantages of CNN in capturing local information and Transformer in modeling global dependencies, thereby effectively improving training speed and overall performance while maintaining model expressibility. The transformer model in CNN-transformer and bert in CNN-BERT model both play the role of self-attention mechanism. However, the parameters of BERT models are large, and transformer provides greater flexibility.

3.4 Data Labeling

The manual labeling process in this study references the 73 AI-related keywords proposed by Yao et al. (2024) as the initial basis for annotation [12]. Building on this foundation, the research team provided systematic training to the annotators and ensured the accuracy and consistency of the labeling process through sample-based learning.

In order to further ensure the accuracy and consistency of the annotations, the method of cross-validation is adopted in this report. The trained taggers were divided into two groups and annotated the same text independently. Only when two groups of taggers have the same tagging results for the same text, the tagging result is considered valid. In the case of annotation differences, the team will conduct in-depth discussion and review until a consensus is reached. In addition, all annotations are ultimately reviewed and confirmed by senior researchers to ensure the overall quality and consistency of the data. Ultimately, 10,000 samples were labeled, the data is partitioned into an 80% training set and a 20% test set.

3.5 Model Training and Evaluation

The training set is used to train the models, and their performance on the test set is assessed. F1-Score, AUC, Accuracy, Recall, and Precision are among the metrics used for assessment. These metrics comprehensively assess the models' performance in classification tasks, ensuring the scientific validity and reliability of the results. This study validates the exceptional performance of the suggested CNN-Transformer model in AI patent identification tasks after analyzing the performance of several models on these criteria.

3.6 Experimental Environment

All computational experiments in this study are conducted on high-performance computing servers equipped with NVIDIA 3060 GPUs to ensure the efficiency of deep learning model training. This study uses Python 3.9.20 as the programming language, TensorFlow 2.10.0 as the deep learning framework, and scikit-learn for traditional machine learning implementation.This setup ensures the efficiency and stability of model training and provides a foundation for the reproducibility of the results.

4. Research Results

4.1 Model Performance Analysis

This research examines the efficacy of algorithms—namely, Decision Tree, Tured Decision Tree, Random Forest, Tured Random Forest, SVM, Tured SVM, CNN, LSTM, BERT, CNN-BERT, and CNN-Transformer—with relation to the classification of patent texts. The main objective is to precisely determine the presence of AI technologies within patents, thereby improving the speed and precision of patent text analysis.

Five metrics—Accuracy, Precision, Recall, F1-score, and the AUC—were used to assess each algorithm's performance in detail.

(1) Accuracy

The ratio of correctly predicted cases to the whole dataset is known as the accuracy measure, and it is calculated using the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. An elevated accuracy score signifies a superior general predictive capability of the model.

(2) Precision

The precision metric assesses the ratio of true positive instances to the total instances labeled as positive, computed by the formula:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

A high precision value signifies a minimal false positive rate, suggesting that the model seldom misclassifies negative instances as positive.

(3) Recall

The recall measure, which is calculated using the following formula, assesses the proportion of correctly detected real positive instances to all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

An elevated recall level indicates a minimal false negative rate, suggesting that the model rarely misses identifying positive instances.

(4) F1-score

The F1-score provides a fair evaluation of the model's accuracy and comprehensiveness since it is the harmonic average of precision and recall. Its calculation is as follows:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

An increased F1-score denotes enhanced performance in both precision and recall.

(5) AUC

The area under the ROC curve, or AUC, indicates how well the model distinguishes between positive and negative results. AUC values that are closer to 1 indicate improved the accuracy of models.

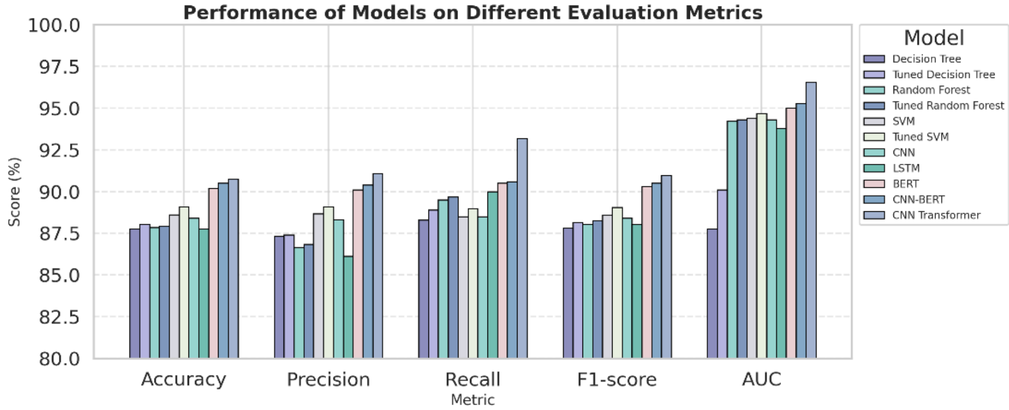

Fig. 2 Model Performance Across Metrics

Fig. 2 illustrates the comparative performance of Decision Tree, Tured Decision Tree, Random Forest, Tured Random Forest, SVM, Tured SVM, CNN, LSTM, BERT, CNN-BERT, and CNN-Transformer in terms of accuracy, precision, recall, F1-score, and AUC.

As shown in the data above, deep learning models generally outperform traditional machine learning models. Among all evaluation metrics, the CNN-Transformer model exhibits superior performance, achieving the highest scores in Accuracy (90.75%), Precision (91.10%), Recall (93.20%), F1-score (90.97%), and AUC (96.56%). This indicates that the CNN-Transformer model significantly outperforms other models in comprehensive classification performance.

The CNN-Transformer model combines the local feature extraction capabilities of CNNs with the attention mechanisms of Transformers, enhancing classification accuracy and robustness by better capturing semantic information in patent texts. The results demonstrate the effectiveness of this integrated approach in handling complex text data, providing strong technical support for patent text analysis.

4.2 AI Patent Analysis

(1) Overview of AI Patents

This study utilizes the trained CNN-Transformer model to classify and identify patents of A-share listed manufacturing enterprises. Fig. 3 illustrates the cumulative number of AI patents and non-AI patents from 2014 to 2023. The x-axis displays the number of years and the y-axis shows the total number of patents.

Green bars in the chart represent AI patents, whereas orange bars represent non-AI patents.

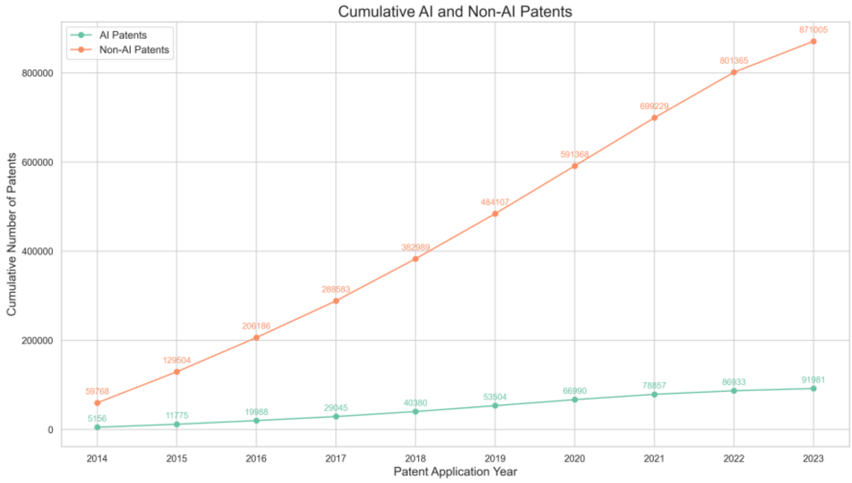

Fig. 3 Total Number of Patents from 2014 to 2023

Fig. 3 shows a steady increase in cumulative patent filings by A-share listed manufacturing enterprises from 2014 to 2023 in both AI and non-AI categories. Non-AI patents grew from 59,768 to 871,005, while AI patents rose from 5,156 to 91,981. This sustained growth highlights continued investment in innovation across the manufacturing sector.

(2) Top Ten Enterprises in AI Patent Applications

This study analyzes the top ten A-share listed manufacturing enterprises by AI patent applications, as shown in Table 4. Gree Electric Appliances leads with 4,871 patents, followed by BOE Technology Group with 4,232. Midea Group and Visionox Holdings hold 3,110 and 3,085 patents, ranking third and fourth. Hikvision and Haier Smart Home follow with 2,936 and 2,919 patents, respectively. ZTE, Dahua Technology, BYD, and Sichuan Changhong have 2,823, 1,767, 1,451, and 1,421 patents, respectively.

Table 3

Total Number of Patents (2014-2023) of Top Ten Enterprises

Stock Code	Company Name	Total Number of AI Patents
000651	Gree Electric Appliances, Inc. of Zhuhai	4,871
000725	BOE Technology Group Co., Ltd.	4,232
000333	Midea Group Co., Ltd.	3,110
002841	Guangzhou Shiyuan Electronic Technology Company Limited	3,085
002415	Hangzhou Hikvision Digital Technology Co.,Ltd.	2,936
600690	Haier Smart Home Co., Ltd.	2,919
000063	ZTE Corporation	2,823
002236	Zhejiang Dahua Technology Co.,Ltd.	1,767
002594	Build Your Dreams Co., Ltd.	1,451

600839	Sichuan Changhong Electric Co., Ltd.	1,421

As shown in Table 3, Gree Electric Appliances leads in AI patent applications, highlighting its strong commitment to AI R&D. BOE Technology Group and Midea Group also show notable competitiveness, reflecting substantial investments in the field.

Fig. 4 further illustrates the AI patent trends of these ten companies from 2014 to 2023. From 2014 to 2019, most experienced rapid growth, especially Gree, BOE, and Midea. Gree’s AI patents rose from 173 in 2014 to 1,062 in 2019, while BOE increased from 208 to 794, reflecting intensified focus on AI innovation during this period.

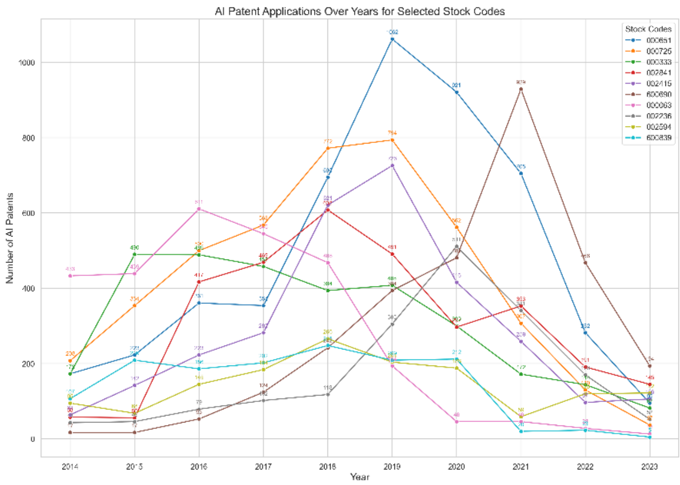

Fig. 4 Annual Patent Trend of Top Ten Enterprises (2014-2023)

As shown in Fig. 4, since 2020, AI patent applications by most enterprises have declined or stabilized. ZTE and Dahua Technology saw sharp drops to 13 and 52 patents, respectively. Haier Smart Home peaked at 929 patents in 2021 but fell to 194 soon after. Midea Group and BYD showed similar downward trends post-2020.

Overall, the ten enterprises show early intensive AI innovation, followed by a decline reflecting technological maturity, strategic shifts, or market changes. This pattern highlights key trends in manufacturing-oriented AI development.

5. Conclusion and Future Work

This study presents a novel deep learning model that integrates CNN with Transformer architectures to advance the identification and classification of AI patents. By integrating CNN's expertise in local feature extraction with the Transformer's ability to capture global context, the proposed CNN-Transformer model effectively tackles the complexities in patent text analysis. Comparative experiments show that CNN-Transformer model is superior to traditional machine learning algorithms, such as decision trees, random forests, SVM, BERT and Bert-CNN models.

The proposed CNN-Transformer hybrid model significantly enhances text analysis for complex, domain-specific content such as patent documents. By integrating CNN for local semantics and Transformer for long-range dependencies, it improves classification accuracy and efficiency, facilitating informed R&D decision-making.

While this study focuses on AI-related patent classification, the proposed model architecture is readily adaptable to various text classification tasks, such as news categorization, academic topic identification, legal document analysis, and sentiment detection. Its ability to process lengthy and domain-specific texts demonstrates strong generalizability and broad applicability in NLP scenarios.

Acknowledgements

This work was supported by the Undergraduate Innovation Program of GDUT (Project No. xj2024118451036, "Research on Business Model Evolution Based on CNN-GRU").

REFERENCES

[1] *B. Li, F. Jiang, H. Xia, and J. Pan*, Under the background of AI application, research on the impact of science and technology innovation and industrial structure upgrading on the sustainable and high-quality development of regional economies, Sustainability**, vol.** 14**,** p. 11331, 2022.

[2] *A. Redchuk and F. Walas Mateo*, New Business Models on Artificial Intelligence—The Case of the Optimization of a Blast Furnace in the Steel Industry by a Machine Learning Solution, Applied System Innovation**, vol.** 5**,** p. 6, 2021.

[3] *A. Korinek and J. E. Stiglitz*, Artificial intelligence, globalization, and strategies for economic development, National Bureau of Economic Research 2021.

[4] *Q. Liu and J. Li*, The progress of business analytics and knowledge management for enterprise performance using artificial intelligence and man-machine coordination, Journal of Global Information Management (JGIM)**, vol.** 30**, pp.** 1-21, 2022.

[5] *Y. Liu, W. Fu and D. Schiller*, The making of government-business relationships through state rescaling: a policy analysis of China's artificial intelligence industry, Eurasian Geography and Economics**, pp.** 1-29, 2024.

[6] *M. Sharma, S. Luthra, S. Joshi, and A. Kumar*, Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy, Government Information Quarterly**, vol.** 39**,** p. 101624, 2022.

[7] *W. Xie, R. Chen and Z. Li*, Leveraging machine learning to uncover the dynamic evolution of business models in intelligent manufacturing, Computers & Industrial Engineering**, vol.** 197**,** p. 110597, 2024.

[8] *J. Liu, Y. Qian, Y. Yang, and Z. Yang*, Can Artificial Intelligence Improve the Energy Efficiency of Manufacturing Companies? Evidence from China, International Journal of Environmental Research and Public Health**, vol.** 19**,** p. 2091, 2022-2-13 2022.

[9] *Q. P. Nguyen and D. H. Vo*, Artificial intelligence and unemployment:An international

evidence, Structural Change and Economic Dynamics, **vol.** 63, pp. 40-55, 2022.

[10] *B. Guo, C. Zhang, J. Liu, and X. Ma*, Improving text classification with weighted word embeddings via a multi-channel TextCNN model, Neurocomputing, **vol.** 363, pp. 366-374, 2019.

[11] *S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M. Shyu, S. Chen, and S. S. Iyengar*, A survey on deep learning: Algorithms, techniques, and applications, ACM computing surveys (CSUR), **vol.** 51, pp. 1-36, 2018.

[12] *G. Graetz and G. Michaels*, Robots at Work, The Review of Economics and Statistics, **vol.** 100, pp. 753-768, 2018.

[13] *R. C. Deo*, Artificial intelligence and machine learning in cardiology, Circulation, **vol.** 149, pp. 1235-1237, 2024.

[14] *K. Govindan*, Unlocking the potential of quality as a core marketing strategy in remanufactured circular products: A machine learning enabled multi-theoretical perspective, International Journal of Production Economics, **vol.** 269, p. 109123, 2024.

[15] *J. Devlin, M. Chang, K. Lee, and K. Toutanova*, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171-4186.

[16] *N. Rane*, ChatGPT and similar Generative Artificial Intelligence (AI) for building and construction industry: Contribution, Opportunities and Challenges of large language Models for Industry 4.0, Industry 5.0, and Society 5.0, Opportunities and Challenges of Large Language Models for Industry, **vol.** 4, 2023.

[17] *Z. Sun, X. Li, X. Sun, Y. Meng, X. Ao, Q. He, F. Wu, and J. Li*, Chinesebert: Chinese pretraining enhanced by glyph and pinyin information, arXiv preprint arXiv:2106.16038, 2021.

[18] *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.

[19] *A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman*, BERT-CNN: A Deep Learning Model for Detecting Emotions from Text., Computers, Materials & Continua, **vol.** 71, 2022.

[20] *Y. Rong, S. Jianfei and Z. Xiaowei*, Research on knowledge distillation method for news text classification based on BERT-CNN, Application of electronic technology, **vol.** 49, pp. 8-13, 2023.

[21] *M. Wankhade, A. C. S. Rao and C. Kulkarni*, A survey on sentiment analysis methods, applications, and challenges, Artificial Intelligence Review, **vol.** 55, pp. 5731-5780, 2022.

[22] *M. Hsu, Y. Hsin and F. Shiue*, Business analytics for corporate risk management and performance improvement, Annals of Operations Research, pp. 1-41, 2022.

[23] *L. Jiang-hua*, Support Vector Machine Training Algorithm: A Review[J]., 2002.

[24] *T. Li, X. Zhang, S. Zhang, and L. Wang*, Self-supervised learning with a dual-branch ResNet for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters, **vol.** 19, pp. 1-5, 2022.

[25] *A. Amro, M. Al-Akhras, K. E. Hindi, M. Habib, and B. A. Shawar*, Instance reduction for avoiding overfitting in decision trees, Journal of Intelligent Systems, **vol.** 30, pp. 438-459, 2021.

[26] *L. Rokach*, Decision forest: Twenty years of research, Information Fusion, **vol.** 27, pp. 111-125, 2016.

[27] *T. He, W. Huang, Y. Qiao, and J. Yao*, Text-attentional convolutional neural network for scene text detection, IEEE transactions on image processing, **vol.** 25, pp. 2529-2541, 2016.

[28] *S. Hochreiter*, Long Short-Term Memory, 1997.