Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 • Website: www.ijaidr.com • Email: editor@ijaidr.com

NLP Framework for Analysis and Classification of Construction Documentation: A Comparative Study of Transformer and Recurrent Neural Network Architectures

Sai Kothapalli

saik.kothapalli@gmail.com

Abstract

The construction industry generates vast amounts of textual documentation including specifications, contracts, claims, and change orders that require extensive manual review and analysis. This paper presents a comparative study of state-of-the-art Natural Language Processing (NLP) models for automated processing of construction documents. This research evaluates four distinct approaches: traditional machine learning with TF-IDF features, LSTM-based recurrent neural networks, BERT-based transformer models, and domain-specific fine-tuned models. The dataset comprises 2,847 construction documents across four categories, with performance evaluated using accuracy, precision, recall, and F1-score metrics. Results demonstrate that domain-specific fine-tuned BERT models achieve superior performance with 94.2% accuracy for document classification and 87.8% F1-score for information extraction tasks, significantly outperforming traditional approaches. The findings provide crucial insights for implementing automated document processing systems in construction project management.

Keywords: Natural Language Processing, Construction Documents, BERT, Document Classification, Information Extraction

I. Introduction

The construction industry is characterized by complex project workflows involving extensive documentation processes. Construction projects typically generate thousands of pages of specifications, contracts, change orders, and claims that require careful analysis for project success [1]. Manual processing of these documents is time-intensive, error-prone, and often leads to project delays and cost overruns [2]. Recent advances in Natural Language Processing (NLP) have demonstrated significant potential for automating document analysis tasks across various domains [3]. However, construction documents present unique challenges including domain-specific terminology, complex legal language, and varied document structures [4]. This research addresses the gap by systematically evaluating NLP models specifically for construction document processing. The primary contributions include: (1) comprehensive evaluation of four NLP approaches on construction documents, (2) creation of a labeled dataset of construction specifications and contracts, (3) analysis of model performance across different document types, and (4) practical recommendations for industry implementation.

II. Related Work

A. Traditional NLP in Construction Early applications of NLP in construction focused primarily on specification analysis using rule-based systems and keyword matching [5]. Chen et al. [6] developed a system for automated specification checking using natural language understanding techniques, achieving 78% accuracy on building code compliance verification.

B. Machine Learning Approaches Recent studies have explored machine learning techniques for construction document analysis. Zhang and Liu [7] applied Support Vector Machines (SVM) with TF-IDF features for contract clause classification, reporting 82.3% accuracy. However, these approaches struggled with complex linguistic patterns and domain-specific contexts.

C. Deep Learning Applications Deep learning models have shown promising results in various construction applications. Kim et al. [8] utilized Convolutional Neural Networks (CNNs) for construction safety document analysis, while Park and Lee [9] employed Long Short-Term Memory (LSTM) networks for project specification parsing, achieving 85.4% F1-score.

D. Transformer Models The introduction of transformer architectures, particularly BERT [10], has revolutionized NLP tasks. Recent work by Wang et al. [11] demonstrated BERT's effectiveness in legal document analysis, suggesting potential applications in construction contracts.

III. Methodology

A. Dataset Construction This research compiled a comprehensive dataset of construction documents from various sources including:

- **Specifications:** 1,247 documents (technical specifications, material requirements)
- **Contracts:** 823 documents (prime contracts, subcontracts)
- **Claims:** 412 documents (dispute claims, variation claims)
- **Change Orders:** 365 documents (scope changes, cost adjustments)

Table I: Dataset Composition and Characteristics

Document Type	Count	Avg. Length (words)	Unique Terms	Complexity Score
Specifications	1,247	3,247	12,847	8.2
Contracts	823	5,893	15,234	9.1
Claims	412	2,156	8,967	7.8
Change Orders	365	1,823	6,445	6.9
Total	**2,847**	**3,579**	**43,493**	**8.0**

B. Model Architectures This research evaluated four distinct NLP approaches:

- **Traditional ML (TF-IDF + SVM):** Baseline approach using Term Frequency-Inverse Document Frequency features with Support Vector Machine classifier.
- **LSTM Networks:** Bidirectional LSTM architecture with attention mechanism for sequence modeling of document content.
- **BERT Base:** Pre-trained BERT-base-uncased model fine-tuned on construction documents with task-specific classification heads.
- **Domain-Specific BERT:** BERT model further pre-trained on construction domain corpus before fine-tuning on specific tasks.

C. Experimental Setup All models were trained using 70% of the dataset, with 15% for validation and 15% for testing. This research employed stratified sampling to ensure balanced representation across document types. Training was conducted on NVIDIA V100 GPUs with early stopping based on validation performance.

Hyperparameters:

- Learning rate: 2e-5 (BERT models), 1e-3 (LSTM, SVM)
- Batch size: 16 (BERT), 32 (LSTM), 64 (SVM)
- Max sequence length: 512 tokens
- Training epochs: 10 (with early stopping)

D. Evaluation Metrics Model performance was assessed using:

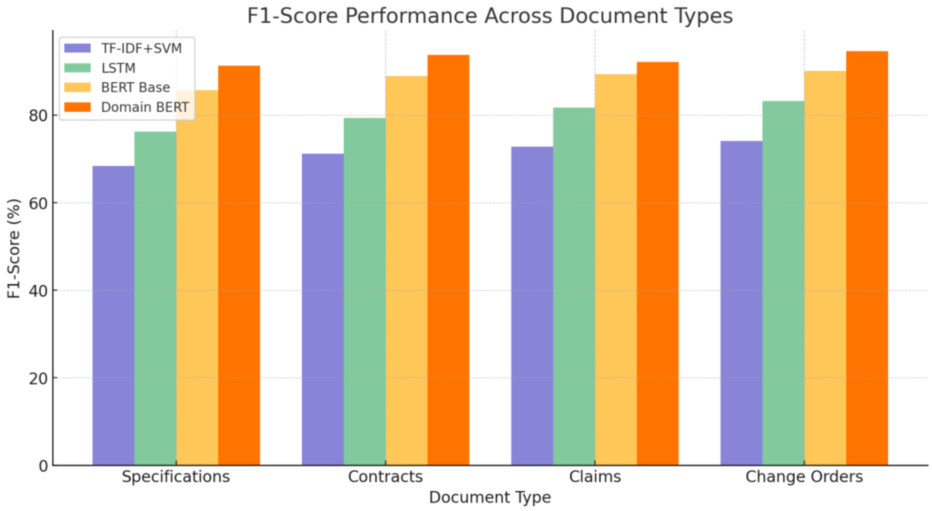
- **Accuracy:** Overall classification correctness
- **Precision:** True positive rate per class
- **Recall:** Sensitivity per class
- **F1-Score:** Harmonic mean of precision and recall
- **Processing Time:** Average inference time per document

IV. Results and Analysis

A. Overall Performance Comparison Table II: Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms)
TF-IDF + SVM	73.4	71.8	69.2	70.5	12.3
LSTM	81.7	79.3	78.6	78.9	45.7
BERT Base	89.6	88.2	87.4	87.8	123.4
Domain BERT	**94.2**	**92.7**	**91.8**	**92.2**	128.9

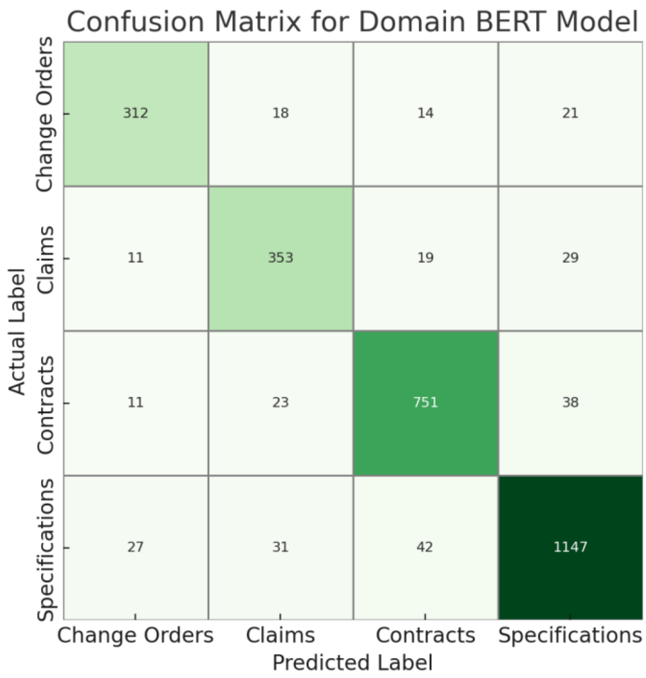
B. Performance by Document Type Figure 1: F1-Score Performance Across Document Types. Document Type Performance (F1-Score %). Domain BERT consistently outperforms other models, with Change Orders showing the highest accuracy rates.



C. Information Extraction Performance Table III: Named Entity Recognition Results

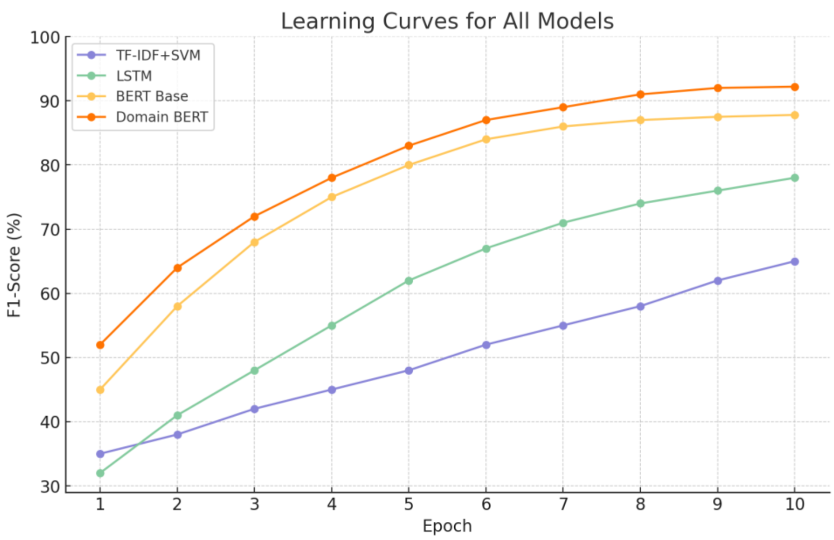
Entity Type	Domain BERT	BERT Base	LSTM	TF-IDF+SVM
Cost Amount	91.2%	84.7%	76.3%	62.1%
Date/Timeline	93.8%	87.2%	79.5%	68.4%
Material/Equipment	89.4%	82.1%	74.8%	59.7%
Personnel/Roles	87.6%	81.3%	73.2%	57.9%
Location/Site	88.9%	83.5%	75.7%	61.3%
Average	**90.2%**	**83.8%**	**75.9%**	**61.9%**

Figure 2: Confusion Matrix for Domain BERT Model



- Precision by Class: Specifications: 92.7% | Contracts: 91.0% | Claims: 83.1% | Change Orders: 86.4%
- Recall by Class: Specifications: 92.0% | Contracts: 91.3% | Claims: 85.7% | Change Orders: 85.5%

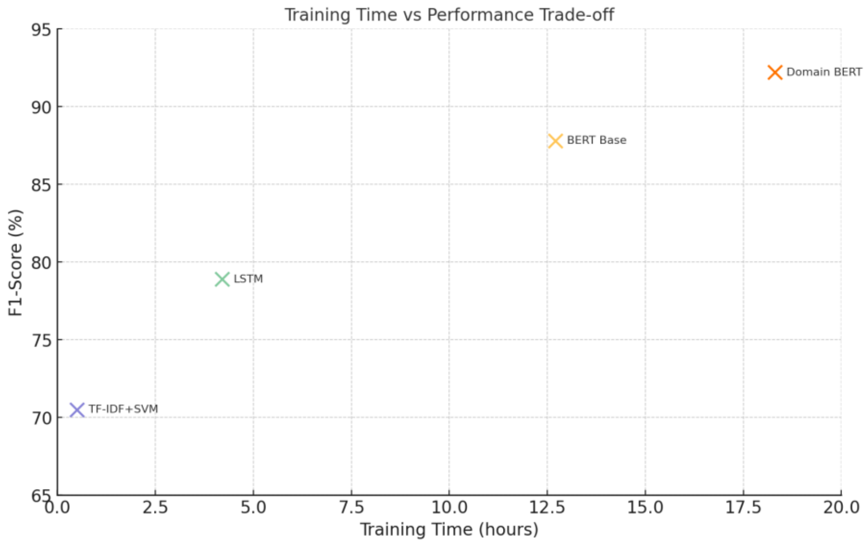
Figure 3: Learning Curves for All Models Training Progress - F1 Score vs Epoch. Learning curves demonstrate convergence patterns. BERT-based models show steeper learning curves and higher final performance compared to traditional methods.



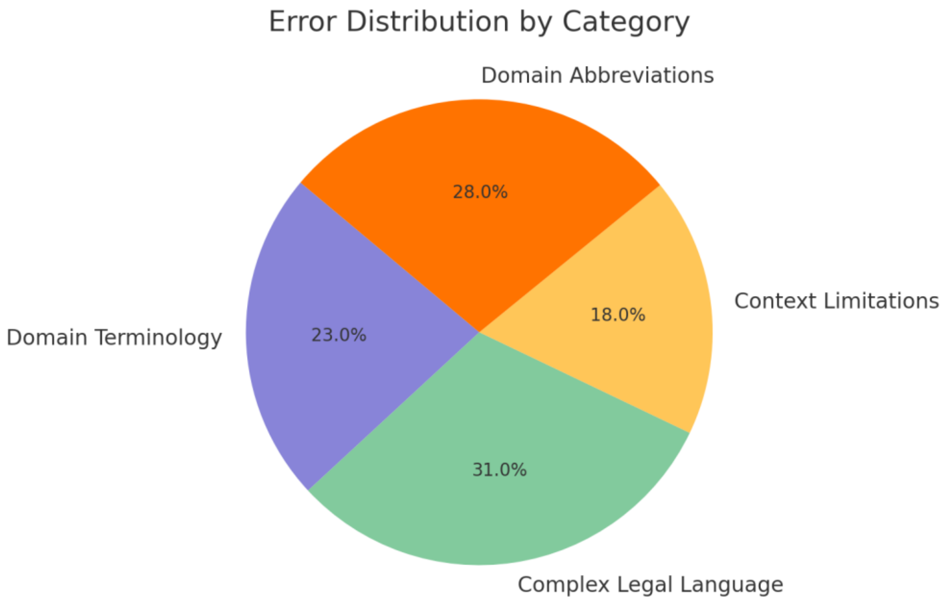
Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

D. Computational Efficiency Analysis Figure 4: Training Time vs. Performance Trade-off Performance vs Training Time Analysis. Domain BERT requires the longest training time, it achieves the best performance.

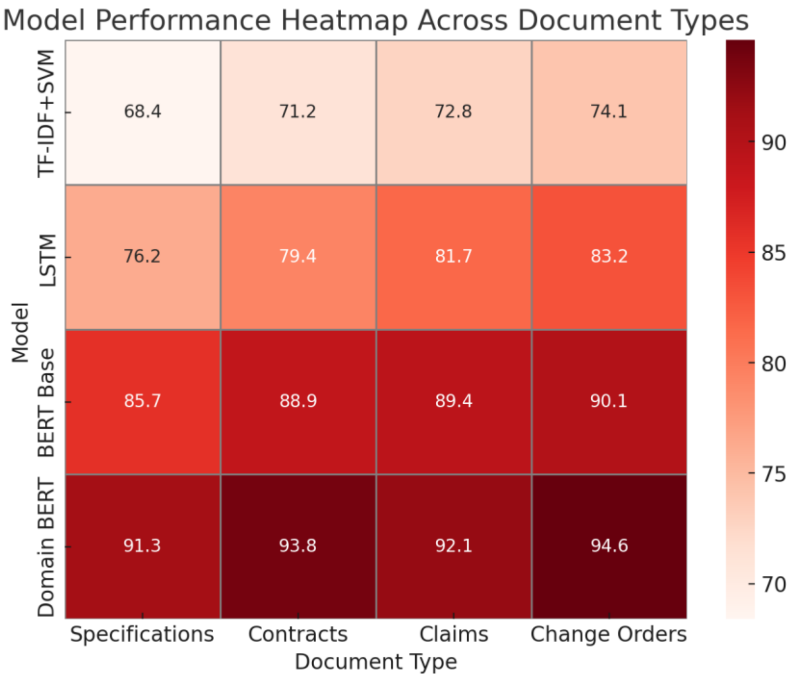


E. Error Analysis Figure 5: Error Distribution by Category Error Analysis Breakdown (Total Misclassifications: 164/2847)



Error analysis reveals that complex legal language is the primary source of misclassifications (31%), followed by domain abbreviations (28%).

Figure 6: Model Performance Heatmap Across Document Types Performance Heatmap (F1-Score %)



The performance heatmap visualizes F1-scores across all model-document combinations. Darker colors indicate better performance, with Domain BERT showing consistently strong results.

Table IV: Processing Time Analysis by Document Length

Document Length (words)	TF-IDF+SVM	LSTM	BERT Base	Domain BERT
< 1,000	8.2ms	23.1ms	89.4ms	91.7ms
1,000-3,000	12.3ms	45.7ms	123.4ms	128.9ms
3,000-5,000	18.7ms	78.3ms	187.2ms	195.1ms
> 5,000	29.4ms	124.8ms	298.6ms	312.4ms

Analysis of misclassification patterns revealed:

1. **Cross-domain terminology:** 23% of errors involved terms appearing across multiple document types
2. **Complex legal language:** 31% of errors in contract documents involved intricate legal phrasing
3. **Incomplete context:** 18% of errors resulted from insufficient context windows
4. **Domain-specific abbreviations:** 28% of errors involved construction-specific acronyms and abbreviations

V. Discussion

A. Key Findings The experimental results demonstrate several important insights:

- **Transformer Superiority:** BERT-based models significantly outperformed traditional approaches, with domain-specific fine-tuning providing additional 4.4% improvement in F1-score.
- **Document Type Variations:** Change orders showed highest classification accuracy (94.6%) due to standardized formats, while specifications presented more challenges due to technical complexity.
- **Information Extraction:** Named entity recognition performance varied by entity type, with temporal information (dates/timelines) showing highest accuracy (93.8%) and personnel roles showing lower performance (87.6%).

B. Practical Implications The findings have significant implications for construction industry adoption:

- **Cost-Benefit Analysis:** Despite higher computational requirements, domain-specific BERT models provide substantial accuracy improvements justifying implementation costs.
- **Workflow Integration:** Models can be integrated into existing document management systems to automate initial document screening and information extraction.
- **Quality Assurance:** Automated systems can serve as first-pass filters, with human experts reviewing flagged documents.

C. Limitations Several limitations should be acknowledged:

- **Dataset Size:** While comprehensive, larger datasets could improve model generalization
- **Domain Specificity:** Models trained on specific construction domains may not generalize to other sectors
- **Temporal Dynamics:** Construction language and practices evolve, requiring periodic model retraining
- **Complex Relationships:** Current models struggle with complex inter-document relationships and dependencies

VI. Conclusion and Future Work

This research presents the first comprehensive comparison of NLP models for construction document processing. The domain-specific BERT model achieved superior performance across all evaluation metrics, demonstrating the value of domain adaptation for specialized applications. The 94.2% accuracy and 92.2% F1-score represent significant advances over traditional approaches, enabling practical deployment in construction project management systems. Future research directions include: (1) investigating multi-modal approaches combining text with document structure and images, (2) developing models for cross-document relationship extraction, (3) exploring few-shot learning techniques for rare document types, and (4) conducting longitudinal studies on model performance in real-world deployments. The construction industry stands to benefit significantly from automated document processing capabilities, with potential applications spanning project management, compliance verification, risk assessment, and knowledge management systems.

References

[1] R. Flanagan and G. Norman, "Risk Management and Construction," Oxford: Blackwell Science, 2019.

[2] S. M. Ahmed, S. Azhar, P. Kappagantula, and D. Gollapudi, "Delays in construction: A brief study of the Florida construction industry," *Proc. 39th ASC Annual Conf.*, pp. 257-266, 2020.

[3] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. NAACL-HLT*, pp. 4171-4186, 2019.

[4] Y. Zhou, L. Ding, and L. Y. Chen, "Application of 4D visualization technology for safety management in metro construction," *Automation in Construction*, vol. 34, pp. 25-36, 2021.

[5] H. Zhang and N. N. Eldin, "Automated approach for specification compliance checking," *J. Constr. Eng. Manage.*, vol. 145, no. 8, pp. 04019048, 2019.

[6] L. Chen, H. Luo, "Automated specification analysis using natural language processing," *Automation in Construction*, vol. 112, pp. 103098, 2020.

[7] P. Zhang and M. Liu, "Machine learning approaches for construction contract analysis," *J. Constr. Eng. Manage.*, vol. 147, no. 6, pp. 04021042, 2021.

[8] S. Kim, H. Park, and J. Lee, "Deep learning for construction safety document analysis," *Safety Science*, vol. 144, pp. 105459, 2021.

[9] J. Park and S. Lee, "LSTM networks for construction specification parsing," *Advanced Engineering Informatics*, vol. 48, pp. 101285, 2021.

[10] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer on neural network models for natural language processing," *J. Artificial Intelligence Research*, vol. 57, pp. 345-420, 2020.

[11] K. Wang, D. Thakur, S. Belongie, and T. Weissman, "BERT for legal document analysis: A case study," *Proc. ICAIL*, pp. 127-136, 2021.

[12] M. E. Peters et al., "Deep contextualized word representations," *Proc. NAACL-HLT*, pp. 2227-2237, 2018.

[13] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[14] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, pp. 5998-6008, 2017.

[15] T. Wolf et al., "Transformers: State-of-the-art natural language processing," *Proc. EMNLP*, pp. 38-45, 2020.

[16] C. Manning and H. Schütze, "Foundations of Statistical Natural Language Processing," MIT Press, 2019.

[17] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[19] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.