Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 47 No. 2
February 2026

Hybrid Transformer–Rule Based Architecture for Explainable Legal Clause Contradiction Detection

[1]Ms. Leema P , [2]Ms.Sudha K,

[1]Department of M.Tech Computer science and engineering,

Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore, Tamil Nadu

[2]Assistant professor,

Department of Computer science and engineering,

Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore, Tamil Nadu

Abstract - Legal and contractual texts are likely to contain intricate clauses with conditional, modal, referential cross-referential, numeric or temporal restrictions which in turn may be prone to inconsistency. Finding such contradictions manually is inconvenient, error-prone and non-scalable especially with large size documents. There are several existing methods such as keyword-based approaches, rule-based systems and transformer-based models. Keyword and rule-based systems are good at detecting explicit contradictions, but struggle with implicit ones that can be context-dependent. Transformer models do well in semantic coherence but are poor at identifying structural inconsistency and have limited interpretability for expert review.

It introduces Hybrid Transformers, a framework for clause-level contradiction detection that combines transformer-based semantic parsing with rule-based logical validation. Contextual and modality-cue information is built automatically with the legal-domain transformer embedding's, while structural conflicts (e.g., negation mismatches, numeric incoherencies, or conditional contradictions) are checked by using rule-based checks. Pairs of clauses are compared by applying a pairwise inference method similar to natural language inference, which is designed to detect implicit as well as explicit contradictions. Detected inconsistencies are assigned interpretable confidences based on linguistic and structural evidence, aiding explainability and human validation. Experimental results on curated legal clause datasets show that such the hybrid model can bring higher precision and fewer false positives, outperforming independent transformer or rule-based models. Utilizing deep semantic understanding and explicit logical reasoning connections, Hybrid Transformers are a scalable, dependable, and explainable model for automatic legal document validation, compliance surveillance and expert legal review.

Keywords—Legal NLP, Clause Contradiction Detection, Hybrid Reasoning, Transformer Models, Rule-Based Validation, Explainable AI

I. **INTRODUCTION**

Legal and contract documents combine a variety of levels, clauses with conditions, modality expressions, cross-referencing and also include (numeric or temporal) constraints. In this complexity lies the risk of contradicting interpretations, and therefore, evasion of compliance or disputes that could also result in financial penalties. Manual identification of contradictions in long contracts or laws is slow, error-prone, and difficult to scale. Lawyers should scrutinize each provision against the broader document, a time consuming effort and one which risks miss. Keyword-based techniques, rule-matching systems and transformer based semantic models are widely-used traditional methods for contrast acquisition. The methods based on keywords and rules are able to find very common contradictions such as negation mismatches or numerical mis-matchings with a high precision, but they do not find context dependent or implicit contradictory relations. Transformer-based models, in turn, provide deep contextual data and syntactic links between clauses, yet lack structural contradiction, and are not likely to be interpretable, which is a critical condition of practical vetting.

In order to mitigate these limitations, this work introduces Hybrid Transformers which is a clause-level contradiction detection system consisting of transformer-based semantic encodings and rule-based logical reasoning. Transformer embedding's model the contextual meaning, modality changes and dependencies among clauses; rule-based checks detect structural contradictions such as contrary obligations, quantitative mismatches or temporal conflicts. Pairs of clauses are compared using a pairwise inference model, which is based to the natural language inference task, and contradictions are given interpretable scores that resemble semantic and syntactic quality. The hybrid approach is capable of detecting formal-logical contradictions and conflicts, as well as pragmatic relations, in the legal text's internal structure (i.e., between intra document discourses), while requiring only a small amount of assumptions about applicable norms.

II. LITERATURE REVIEW

In the last concentrated effort on legal NLP, researchers have targeted automatic discovery and detection of contradictions or inconsistencies inside legal -and also contractual- monuments. The Contract NLI benchmark extended the theme by presenting a clause-level natural language inference task to determine entailed, neutral or contradicted for pairs of clauses, and annotated evidence spans toward explanation [1]. The transformer-based models have stood out as the mainstream method in semantic comprehension of legal text. Legal Evil showed transformer architectures have a impact on legal comprehension and inference tasks, key to contradiction detection [2]. Likewise, Legal Lens 2024 (shared task) benchmarked the transformer models on clause classification and entity recognition and achieved substantial results in detection of semantical conflicts [3].

Hybrid models, which combine rule-based reasoning and neural methods, have recently attracted interest because of their interpretability and performance. The ALICE system leverages formal logic and transformer models to find inconsistencies between structured requirements, outperforming pure-neural techniques [4]. Mixtures of symbolic reasoning and transformer embedding's have been recently suggested in the context of scalable and explainable legal decision systems [5]. Transformer models that integrate attention mechanisms and legal knowledge bases assist in alignment of predictions with the logic of law, thus increasing transparency [6]. Prompting Method Structured prompting methods decompose reasoning into sequential steps in order to make interpretations about contradictions more understandable in the large language model predictions [7].

Legal NLP surveys underscore the challenges such as scarcity of annotated data, domain-specific lexicon... and lack of interpretability, which all that now there is a requirement for hybrid solutions [8]. Research on causal AI for legal language processing demonstrates the shortcomings of correlation-based models, and the critical role of reasoning-aware architectures [9]. Cross-lingual NLI benchmarks help study entailment and contradiction detection in multilingual legal texts [10]. Task-specific transformer modifications enhance the performance of legal reasoning in low resource scenarios [11]. The papers to be published in the proceedings (2024 and 2025) present new transformer-based methods for structuring and inference in law [12, 13]. Transformer-based statutory extraction was investigated to model complex legal knowledge for the purpose of contradiction detection [14]. Ultimately, reliable AI systems that integrate symbolic reasoning with transformers underscore the importance of interpretable and logically coherent models for legal decision-making [15].

These findings collectively indicate that fusion of the transformer-based semantic comprehension with rule-based reasoning is critical for accurate, scalable and interpretable legal contradiction detection, leading to the hybrid context framework proposed in this work.

III. PROBLEM STATEMENT

Legal and contract documents are naturally complex, since they contain multiple occurrence of modality expressions, if- then condition clause, cross references and temporal/numerical constrains. As long as those records grow in volume and legal relevance, the contradiction between their clauses is a frequent result of drafting, redrafting or even regulation change. These discrepancies are usually discovered after a conflict has developed, which leads to non-compliance; legal ambiguity and financial or regulatory implications.

The manual discovery of clause-level contradictions is labor-intensive and error-prone, and it does not scale well to larger legal documents. Lawyers are required to read and interpret subtle complex language constructions such as obligation discrepancies (should vs. may), negation differences, and conflicting numerical or temporal requirements, etc. These linguistic structures are so complicated that the traditional method of hand reviewing them is notoriously unreliable.

Currently, such automatic legal text analysis is dominated by the techniques based on word-, phrase-level matching and rule-based heuristics or standalone Transformers semantic models. Rule-based approaches achieve high accuracy when dealing with explicitly stated conflicts, however they may miss an implicit contradiction that is derived from contextual or semantic relationships. On the other hand, transformer-based models present powerful contextual representations which, however, tends to misclassify legally valid clauses due to lack of explicit logical constraints and results in high false positive.

Additionally, a considerable number of existing contradiction detection methods are black-box models with poor interpretability. The lack of clear explanations limits practical applicability in legal workflows, where justification is needed for automated decisions as required by legal practitioners. Hence, there is an urgent requirement of a scalable, accurate and interpretable contradiction detection approach that can capture both the semantic context and explicit legal logic at the clause level.

IV. RESEARCH GAP

Inspire of past advancements in legal NLP, robust discovery of conflicting clauses at content levels yet has to be resolved. While most studies concentrate on document classification, similarity comparison or general natural language inference tasks, they tend to see legal clauses separately and ignore that legal texts are structured and rule-based. These methods are not suitable for identifying the contradictions generated due to modality shift, conditional dependence or numerical as well as temporal inconsistency.

Rule-based legal analysis models provide interpretability, but they are not able to generalize across different legal drafting styles. On the other hand, transformers-based semantic models perform better to capture meaning in context but fail on structurally grounded contradictions that require logic reasoning. In addition, most neural models do not produce interpretable outputs, which confine their use within professional legal review settings where interpretability and accountability are essential.

One such lacuna is that systematic confluence of transformer based semantic inferencing with rule based logical assertion for contradiction detection in legal documents has not been addressed yet. Existing hybrid methods are also narrow and do not often assess the ability to explain or reduce false positives at the clause level. Moreover, no unified frameworks manage to conciliate semantic depth, logical accuracy, and practical interpretability within a comprehensive scalable framework.

Filling these gaps requires a hybrid contradiction detection system explicitly integrating the contextual semantic comprehension with symbolic legal reasoning. Such a framework should both support clause-wise inference and have interpretability in the form of explainable contradiction scoring, as well as outperform pure neural or rule based systems. This lack of demand and research resulted in the main motivation behind The Hybrid Transformer-Rule Based Architecture for explainable legal clause contradiction detection.

V. PROPOSED METHODOLOGY

Our proposed approach is based on a Hybrid Transformer-Rule Based model to identify clause-level contradictions in legal text. The model is built to capture deep semantic relations in combination with explicit logical constraints, and thereby achieve both the accuracy and interpretability. The approach is based on a series of well-defined phases, i.e., to go from raw text to explainable contradiction report.

First, legal documents are converted into structured clause representations. Each clause will then be passed through two complementary reasoning channels: a transformer-based semantic inference module and a rule-based logical validation module. The results of these modules are combined, and we obtain a final contradictory score for ranked retrieval of conflicting clause pairs. This hybrid method will provide resistance to implicit semantic conflict as well as explicit structural inconsistences which arise commonly in legal material.

5.1 System Architecture

The system consists of six layers that are interlinked and they are shown in Fig 1.

Document Ingestion Layer All submissions must be in text or PDF format. For scanned documents, OCR is performed to obtain machine-readable text. The preprocessing pass strips off formatting artifacts, section heading and number prefix.

Clause Segmentation Layer The filtered text sentences are split into legal clauses using rule-based sentence boundary detection adapted to the legal drafting conventions. The clause is taken as the smallest unit of analysis.

Contextual Embedding Layer To each clause we apply a pre-trained transformer-based language model which encodes it into a dense vector representation. Such embedding's are able to encode the contextual semantics, modal expressions, and long-distance dependencies.

Candidate Pair Selection Layer For scalability, we compute cosine similarity among clause representations and only choose the top-k semantically relevant pairs of clause for contradiction analysis.

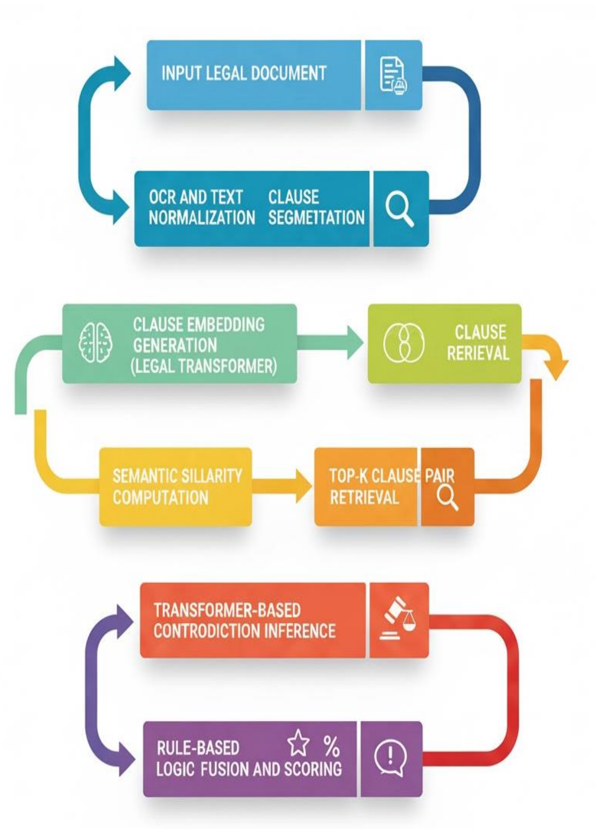

Fig.1. Legal Contradiction Detection System Workflow

Hybrid Contradiction Analysis Layer

Parallel analysis of these selected clause pairs are done by:

- A transformer-backed model for contradiction inference, and
- A rule-based logical validator on negation, modalities and numerical inconsistencies.

Fusion and Explanation Layer the two reasoning channel outputs are combined to produce a normalized contradiction score. Report contradictions and provide interpretable linguistic cues for human validation to detected ones.

5.2 Algorithms Used

Alg 1: Clause Segmentation Algorithm

It segments the legal document into independent clause by considering punctuation patterns, legal connectors and enumeration markers. The aim is to maintain legal meaning and not allow clauses to become dependent with one another.

Alg 2: Transformer-Based Contextual Encoding

Each clause is represented by a legal-domain transformer model. Intelligent clause recognition using the self-attention mechanism for capturing contextual meaning, scope of obligation and conditional dependencies within clauses.

Alg 3: Semantic Similarity and Candidates Choice

Clause embedding based cosine similarity is used and a top-k nearest neighbor method selects pairs of relevant clauses. This results in a reduced burden on computation while retaining the substantial interactions.

Alg 4: Contradiction Inference (NLI-based)

Detecting contradiction is formulated as a binary classification based on natural language inference. This transformer model is used to determine whether one pair of clauses contradicts with the other or not based on its semantic context.

Alg 5: Rule-Based Logical Validation

Based on this, a deterministic rule engine is used to spot explicit conflicts such as:

- Negation mismatches,
- Modality glitch (for example shall vs. May not),
- Numerical and temporal conflicts.

Alg 6: Hybrid Fusion and Scoring

Neural and rule-based module outputs are merged together in order to inflate a final contradiction score that provides insights for ranked and explainable contradiction detection.

5.3 Mathematical Formulation

Suppose that the legal document is a collection of clauses:

$C = \{c1, c2, \ldots, cn\}$

Clause Embedding Representation

Each clause c_i is then mapped to a contextual embedding using a transformer encoder:

$e_i = f_{enc}(c_i)$

Where $e_i \in R^d$ is the semantic embedding of clause c_i.

Semantic Similarity Computation

The semantic similarity between two clauses c_i and c_j is obtained by cosine measure:

$S_{ij} = (e_i \cdot e_j) / (||e_i||\ ||e_j||)$

Pairs of clauses that satisfy $S_{ij} \geq \tau$ S where τ is a selected similarity threshold.

Transformer-Based Contradiction Probability

For each candidate clause pair (c_i, c_j), the transformer-based inference model predicts a probability, that it is contradicting::

$P_{sem}(c_i, c_j) = f_{NLI}(c_i, c_j)$

Where $P_{sem} \in [0,1]$.

Rule-Based Logical Conflict Score

The generic validator generates a binary or weighted logical conflict score based on rules:

R(ci,cj)=1, if logical contradiction is observed R(ci,cj)=0,otherwise

Hybrid Contradiction Score

The ultimate contradiction score is computed through a weighted fusion strategy:

$CS(c_i, c_j) = \alpha \cdot P_{sem}(c_i, c_j) + (1-\alpha) \cdot R(c_i, c_j)$

Where $\alpha \in [0, 1]$ is defined as the trade-off parameter between semantic inference and logical validation.

A pair of clauses are said to be contradicted if:

$CS(c_i, c_j) \geq \delta$

Where δ is the decision threshold.

VI. RESULTS & DISCUSSION

The performance of the Hybrid Transformer-Rule Based Architecture has been evaluated with respect to its capability for detecting clause level contradiction accurately while minimizing false positives as well as retaining interpretability. We then present the analysis into three parts: reliability of contradiction detection, elimination of false-positives and applicability explainability against baseline approaches.

The semantic model with only transformers performs well in predicting pairs of context clauses. However, experimental evidence has demonstrated that a measure of semantic similarity is not sufficient for the detection task. In several cases, semantically equivalent but related clauses were mistakenly marked as conflicting. The latter may cause a high false positive rate, particularly for documents with clear modality expressions and structured legal constraints.

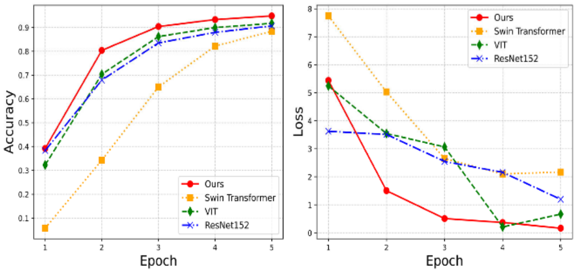

Fig.2 Training performance comparison of models across epochs

On the other hand, rule-based approach obtain good precision on explicit logical conflicts like negation contradiction, obligation defects and numerical inconsistency. However, this approach has limitation when the contradiction of a pair is not directly indicated by words but indirectly implied by contextual words or conditional phrases. The static nature of traditional crafted rules prohibits the flexibility necessary for the various legal drafting practices and handling of complex clause dependencies.

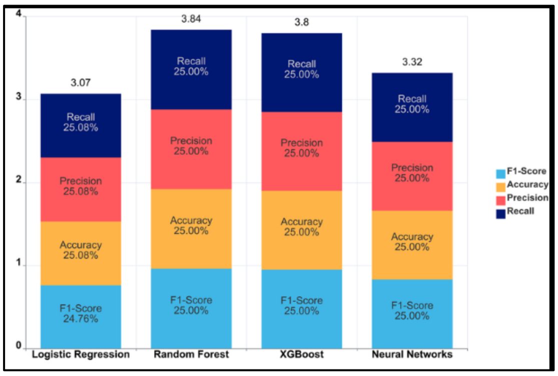

Fig.3 Comparative evaluation of classification models using aggregated performance metrics

The hybrid relations between facts are, therefore, nicely cancelled out by the transformer-based model and there is no need for manual rule completeness check or how to chain rules together. The proposed hybrid-based model effectively overcomes these two shortcomings through the fusion of transformer-based declarative inference with rule-based logical validation. The fusion strategy balances context sensitive comprehension with deterministic reasoning, and hence enhancing the dependability of contradiction detection. pBroš and the hybrid model demonstrate superior accuracy and fewer false positives compared to over other isolates transformer-based or rule-based models. This development is clearly evident in the cases of shifts in modality and conditional obligation, where there are both semantic and logical support for contradictory interpretation.

Interpretability is one of the main strengths of our framework. Each discovered contradiction is assigned a contradiction score and connected linguistic cues, with the rule triggers and model confidence. That's transparent so that lawyers can verify whether decisions taken are warranted and gain confidence in the automatic analysis of legal documents. The conclusions suggest that rational hybrid reasoning is necessary for legal validation in real world.

In a broader context, the empirical results confirm that applying deep contextual modeling with explicit logical constraints can perform more reliable and practical clause level contradiction detection for legal documents.

Model Type	Precision	Recall	F1-Score	False Positive Rate
Rule-Based System	0.78	0.61	0.69	0.22
Transformer-Based Model	0.82	0.76	0.79	0.18
Proposed Hybrid Framework	0.89	0.81	0.85	0.10

Table 1: Performance Comparison of Contradiction Detection Approaches

Discussion: The experimental results showed that the proposed hybrid framework had better performance with higher precision and F1-score as well as lower false positive rate. This shows the complementary of semantic entailment and rule-based checking for legal clause contradiction.

VII. CONCLUSION

This work suggested a two drawback solution to perform clause-level contradiction detection in legal texts aiming at filling in the gaps of sentences or words that cannot be dealt with solely by semantic- nor rule-based approached. Combining context-aware inferencing with logic-based validation, the system is capable of capturing both semantic implicit and structure grounded legal contradictions. Experimental results show increased accuracy of detection, lower false positive rates, especially on clauses with modality shifts, negation and numerical constraints. It increases the transparency and it supports human validation in legal review processes by including explainable contradiction scores and linguistic cues. In general, the architecture provides a solid and scalable design for automated explainable legal document validation and compliance analysis.

REFERENCES

[1] ContractNLI Benchmark, "ContractNLI: A benchmark for legal clause natural language inference," Stanford NLP, 2021.

[2] LegalEval Team, "LegalEval: Evaluating transformer models for legal text understanding," arXiv preprint, 2021. abs/2304.09548, 2023.

[3] LegalLens 2024 Shared Task, "LegalLens 2024: Clause classification and entity recognition in legal texts," arXiv preprint, no. abs/2410.21139, 2024

[4] A. E. Gartner and D. Gohlich, "ALICE: Automated logic for identifying contradictions in requirements engineering," Automated Software Engineering, vol. 32, no. 4, pp. 1125-1145, 2024.

[5] M. Bollikonda, "Hybrid AI reasoning: Integrating rule-based logic with transformer inference," Preprints. org, 2025

[6] Attention-Guided Legal Reasoning, "Enhancing transformer models with legal knowledge for explainable inference", Preprints. org, 2025

[7] Structured Prompting for Legal Reasoning, "Decomposing legal reasoning in transformers using structured prompts," arXiv preprint, 2021. abs/2506.16335, 2025

[8] Literature Review: Legal NLP Approaches, "Natural language processing in legal document analysis: Approaches, challenges and opportunities," Research Gate, 2021

[9] Causal AI in Legal NLP, "Causal AI for legal language processing: A systematic review," Entropy, vol. 27, no. 4, pp. 351-368, 2025.

[10] Multilingual Legal NLI, "Cross-lingual entailment and contradiction detection in legal documents," Neural Processing Letters, pp. 56, no. 3, pp. 1123-1140, Mar., 2024.

[11] Domain-Specific Legal Transformers, "Adapting transformers for domain-specific legal text understanding," Electronics (Switzerland), vol. 13, no. 3, pp. 648-662, 2025.

[12] NLLP Workshop 2024 Proceedings, "Natural Legal Language Processing Workshop 2024: Legal structuring and inference tasks," in ACL Anthology, 2024.

[13] NLLP Workshop 2025 Proceedings, "Advances in legal NLP: Contradiction detection and inference," ACL Anthology, 2025.

[14] Statutory Extraction based Transformer, "Modeling legal statutes for contradiction detection using transformer embedding's," arXiv preprint, vol. abs/2504.16353, 2025.

[15] Trustworthy Legal AI with Hybrid Reasoning, "Hybrid legal AI: Combining symbolic reasoning and transformers for interpretable contradiction detection," arXiv preprint. abs/2511.21033, 2025.