INTERNATIONAL JOURNAL OF INFORMATIVE & FUTURISTIC RESEARCH (IJIFR)
ISSN: 2347-1697
VOLUME 13||ISSUE 8|| APRIL 2026
An Open Access, Peer Reviewed, Cross Reference DOI & Google Scholar Indexed International Research Journal
UGC Approved || IJIFR Impact Factor= 8.057
Estd Year: 2013 || Email: editors@ijifr.org
Availaible Online at: https://www.ijifr.org/

DOI: https://doi.org/10.64672/IJIFR/26.04.13.08.026 PUBLISHED ON: APRIL 18, 2026

SMART AI-BASED FRAMEWORK FOR AUTOMATED LEGAL CLAUSE RISK ASSESSMENT IN CONTRACTS

Bonasi Anusha [1], Usha Rani [2]
[1]M.C.A. Student, [2] Professor
[1,2,] Department of Computer Applications,
Viswam Engineering College, Madanapalle, Andhra Pradesh, India

ABSTRACT

In the modern legal and corporate environment, contract review remains a critical yet resource-intensive task, often requiring significant time, expertise, and manual effort. Traditional approaches to reviewing legal documents are prone to human error, inconsistencies, and inefficiencies, especially when dealing with large volumes of contracts. This paper presents ContractRisk Analyzer AI, an intelligent system designed to automate the identification and classification of legal risk within contract documents using Natural Language Processing (NLP) techniques. The proposed system processes legal contracts in both PDF and plain-text formats, extracting and segmenting the content into individual clauses using advanced sentence tokenization methods. Each clause is evaluated against a structured legal risk knowledge base categorized into High, Medium, and Low risk levels. The classification process is driven by a rule-based NLP engine that ensures transparency by identifying the exact phrases responsible for risk detection. A weighted scoring mechanism aggregates clause-level risks to generate an overall document risk percentage, providing a clear and interpretable assessment. Furthermore, the system integrates an intelligent recommendation module that offers actionable insights based on detected risks, assisting users in making informed legal decisions. Implemented as a web-based application using the Flask framework, the system provides an intuitive user interface with interactive visualizations, including a dynamic risk gauge and clause-level analysis tables. The proposed solution demonstrates the effectiveness of explainable AI in the legal domain, offering a scalable, accessible, and efficient alternative to traditional contract review processes. It significantly reduces review time while maintaining analytical accuracy, making it valuable for legal professionals, organizations, and academic users.

KEYWORDS: ***Contract Analysis; Natural Language Processing; Legal Risk Assessment; Clause Classification; Explainable AI***

PAPER CITATION:
Anusha, B., Rani, U.:"Smart AI-Based Framework for Automated Legal Clause Risk Assessment in Contracts", International Journal of Informative & Futuristic Research (IJIFR), Vol. (13) (8), April 2026, pp. 1099-1107.IJIFR paper ID: 2026/04/IJIFR /V13/E8/026
https://doi.org/10.64672/IJIFR/26.04.13.08.026

1. INTRODUCTION

Legal contracts serve as the backbone of formal agreements across industries, governing obligations, responsibilities, and risk allocation among involved parties. Despite their importance, the manual review of contracts remains a complex, time-consuming, and error-prone process. Legal professionals must carefully analyze multiple clauses, identify potential risks, and ensure compliance with regulatory and organizational standards. This process becomes increasingly challenging with the growing volume and complexity of contractual documents in modern business environments.

The emergence of Natural Language Processing (NLP) has introduced new opportunities for automating document analysis tasks. NLP techniques enable machines to interpret, process, and analyze textual data efficiently, making them suitable for applications in the legal domain. However, many existing automated solutions rely on complex machine learning models that lack transparency and require significant computational resources.

To address these challenges, this paper introduces ContractRisk Analyzer AI, an explainable and lightweight NLP-based system designed to automate the initial screening of legal contracts. The system focuses on identifying risk-related clauses using a rule-based approach, ensuring that every classification is transparent and traceable. By combining efficiency, accessibility, and explainability, the proposed system aims to bridge the gap between traditional legal review practices and modern intelligent automation.

2. LITERATURE SURVEY

The field of automated contract analysis has gained significant attention with the advancement of Natural Language Processing (NLP) and Artificial Intelligence (AI). Researchers and industry practitioners have explored various approaches to improve the efficiency and accuracy of legal document analysis.

Traditional contract review methods rely heavily on manual evaluation performed by legal professionals. While this approach ensures contextual understanding and legal accuracy, it is time-consuming, expensive, and prone to human error, particularly when dealing with large volumes of documents. Studies have highlighted that manual review processes often lead to inconsistencies and missed risk clauses due to fatigue and cognitive overload.

In recent years, machine learning-based approaches have been introduced to automate contract analysis. Models such as Bidirectional Encoder Representations from Transformers (BERT) and domain-specific variants like Legal-BERT have demonstrated strong performance in clause classification and entity extraction tasks. The introduction of datasets such as the Contract Understanding Atticus Dataset (CUAD) has further accelerated research in this domain by providing labeled contract data for training and evaluation. However, these approaches require large datasets, high computational resources, and technical expertise, making them less accessible for practical deployment in smaller organizations.

Commercial solutions such as DocuSign CLM, Ironclad, and Kira Systems offer advanced contract lifecycle management features, including automated clause extraction and risk analysis. Although these platforms provide high accuracy and comprehensive functionalities, they are often associated with high subscription costs and complex deployment requirements, limiting their accessibility to large enterprises.

Rule-based NLP approaches have emerged as a practical alternative, particularly in scenarios where transparency and explainability are critical. These systems rely on predefined keyword patterns and domain-specific knowledge bases to identify and classify risks within text. While they may not capture deep semantic relationships, they offer advantages such as low computational cost, ease of implementation, and complete interpretability of results.

The proposed ContractRisk Analyzer AI system builds upon this rule-based paradigm, focusing on explainable risk classification using a curated legal keyword dictionary. Unlike black-box machine learning models, the system provides explicit reasoning for each classification by highlighting the exact

phrases responsible for risk detection. This makes it particularly suitable for legal applications where justification and traceability are essential.

3. METHODOLOGY

The methodology of the proposed system is designed as a structured pipeline consisting of multiple stages, each responsible for transforming the input contract into a meaningful risk assessment output.

3.1 Data Acquisition and Input Processing

The system accepts legal documents in both PDF and plain-text formats. For PDF files, text extraction is performed using PyMuPDF, ensuring accurate retrieval of content while preserving logical reading order. Plain-text files are decoded using UTF-8 encoding.

3.2 Text Preprocessing

The extracted text undergoes preprocessing to ensure consistency and accuracy in analysis. This includes:

- Removal of extra whitespace and formatting inconsistencies
- Conversion to lowercase for case-insensitive matching
- Sentence tokenization using NLTK to divide the document into clauses

3.3 Risk Classification

Each clause is evaluated against a predefined legal risk dictionary containing categorized keywords:

- **High Risk** (e.g., indemnification, termination without notice)
- **Medium Risk** (e.g., arbitration, limited warranty)
- **Low Risk** (e.g., confidentiality, governing law)

A priority-based matching algorithm ensures that clauses containing multiple risk indicators are classified under the highest applicable risk category.

3.4 Risk Scoring Mechanism

A weighted scoring system is applied to compute the overall document risk:

- High Risk = 3 points
- Medium Risk = 2 points
- Low Risk = 1 point

The final risk score is normalized to a percentage scale (0–100), providing an intuitive measure of overall contractual risk.

3.5 Recommendation Generation

Based on detected risk patterns, the system generates contextual recommendations. These recommendations assist users in understanding potential issues and suggest actions such as

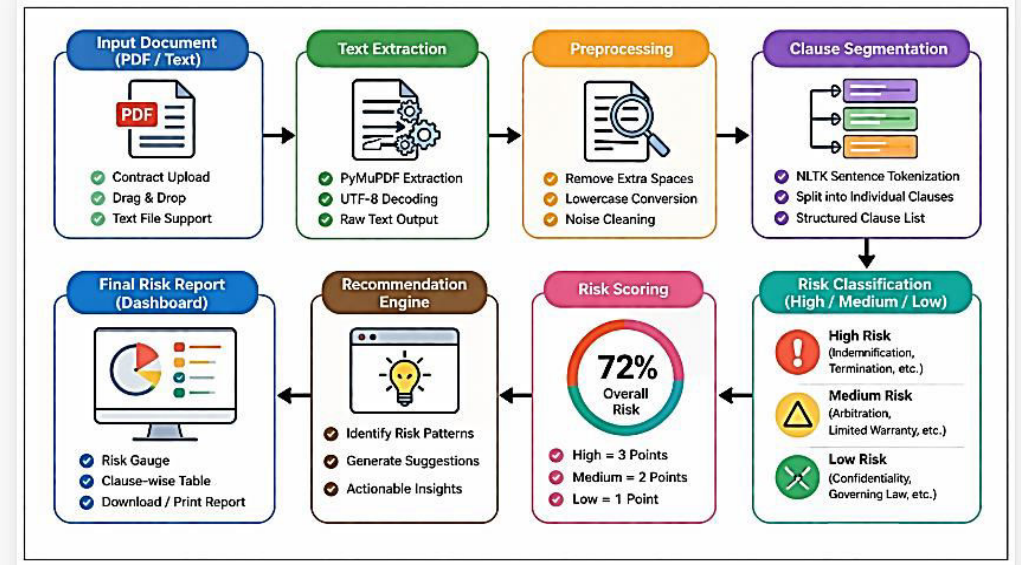

Figure 1: Workflow of ContractRisk Analyzer AI **system**

4. SYSTEM ARCHITECTURE

The architecture of the proposed ContractRisk Analyzer AI system is designed as a modular and layered pipeline to ensure scalability, maintainability, and efficiency. Each layer performs a specific function, transforming the input data into meaningful analytical output.

The system follows six-layer architecture, enabling clear separation of concerns and efficient data processing.

4.1 Architecture Overview

The overall system architecture consists of the following layers:

1. Input Layer
2. Preprocessing Layer
3. Knowledge Base Layer
4. Risk Classification Layer
5. Scoring & Recommendation Layer
6. Presentation Layer

The workflow begins with document ingestion and ends with a comprehensive risk analysis report.

4.2 Layer-wise Description

4.2.1 Input Layer

The Input Layer is responsible for accepting contract documents in multiple formats such as PDF and plain text. The system utilizes PyMuPDF for extracting text from PDF documents and UTF-8 decoding for text files. This ensures flexibility in handling real-world contract formats.

4.2.2 Preprocessing Layer

This layer prepares the raw text for analysis. It performs:

- Text cleaning and normalization
- Conversion to lowercase
- Removal of unwanted whitespace
- Sentence tokenization using NLTK

The output is a structured list of clauses that serve as the fundamental units for risk evaluation.

4.2.3 Knowledge Base Layer

The knowledge base consists of a curated dictionary of legal risk keywords categorized into:

- High Risk
- Medium Risk
- Low Risk

This layer acts as the core intelligence of the system, embedding domain-specific legal knowledge into the analysis process.

4.2.4 Risk Classification Layer

Each clause is analyzed using a rule-based matching algorithm. The system checks for the presence of predefined keywords and assigns a risk level accordingly. A priority mechanism ensures that the highest risk level is assigned when multiple keywords are detected.

4.2.5 Scoring and Recommendation Layer

This layer calculates the overall document risk score using a weighted approach:

- High Risk = 3
- Medium Risk = 2
- Low Risk = 1

The system also generates contextual recommendations based on detected risk patterns, providing actionable insights to users.

4.2.6 Presentation Layer

The Presentation Layer represents the final stage of the system, where the processed results are delivered to the user through an interactive and user-friendly web interface. This layer is designed to transform complex analytical outputs into visually intuitive and easily interpretable information.

It includes the following key components:

- **Risk Score Visualization**:The overall risk score of the document is displayed using a dynamic graphical representation (such as a circular gauge or progress indicator), allowing users to quickly understand the severity of the contract at a glance.
- **Clause-Level Breakdown**:Each clause is presented individually along with its corresponding risk classification and triggering keywords. This enables users to examine specific sections of the contract and identify potential issues in detail.
- **Risk Classification Summary**:A summarized view of the number of High, Medium, and Low risk clauses is provided, offering a clear statistical overview of the document's risk distribution.
- **Recommendations**:Based on the identified risk patterns, the system generates actionable recommendations to guide users in reviewing, modifying, or negotiating specific clauses.

Additionally, the interface supports interactive features such as filtering clauses based on risk levels and exporting the analysis report for documentation purposes. The design emphasizes clarity, responsiveness, and accessibility, ensuring that both technical and non-technical users can effectively interpret the results.

Overall, the Presentation Layer enhances usability by bridging the gap between raw analytical data and meaningful decision-making insights.

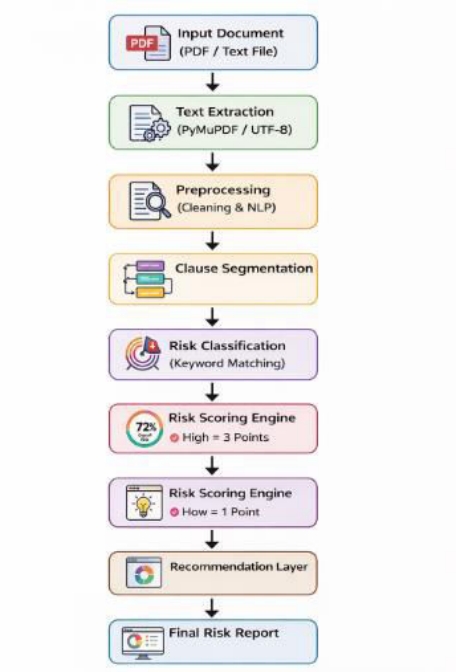

Figure 2: System Architecture of ContractRisk Analyzer AI system

4.3 System Design Advantages

The proposed architecture offers several advantages:

- Modularity**:** Each component can be updated independently
- Scalability**:** Supports extension with additional NLP models
- Explainability**:** Transparent rule-based decision-making
- Efficiency**:** Fast processing without heavy computation
- Accessibility**:** Deployable on low-resource systems

5. IMPLEMENTATION

The implementation of the proposed **ContractRisk Analyzer AI** system is carried out using a modular and scalable approach. The system integrates Natural Language Processing (NLP) techniques with a web-based framework to deliver an efficient and user-friendly contract risk analysis solution.

5.1 System Modules

The system is divided into the following core modules:

1. Document Processing Module
2. Text Preprocessing Module
3. Risk Classification Module
4. Risk Scoring Module
5. Recommendation Engine
6. Web Interface Module

Each module is designed to perform a specific function within the overall processing pipeline.

5.2 Document Processing Module

This module handles the ingestion of input documents. The system supports both:

- PDF files (processed using PyMuPDF)
- Plain-text files (processed using UTF-8 decoding)

The extracted content is converted into a raw text format, ensuring compatibility with subsequent NLP operations.

5.3 Text Preprocessing Module

The preprocessing stage plays a crucial role in improving the accuracy of the system. It involves:

- Removal of redundant spaces and formatting inconsistencies
- Conversion of text into lowercase for uniform processing
- Sentence tokenization using NLTK

This module transforms unstructured text into structured clauses suitable for analysis.

5.4 Risk Classification Module

The classification module is the core analytical component of the system. It uses a rule-based NLP approach to identify risk levels within each clause.

The classification logic follows a priority-based structure:

- If a clause contains High Risk keywords → classified as HIGH
- Else if it contains Medium Risk keywords → classified as MEDIUM
- Else → classified as LOW

This ensures that the most critical risk is always captured.

Algorithm 1:
Input: Clause (sentence)
Output: Risk Level (HIGH / MEDIUM / LOW)
Step 1: For each phrase in High Risk list
If phrase found in clause
Return HIGH
Step 2: For each phrase in Medium Risk list
If phrase found in clause
Return MEDIUM
Step 3: For each phrase in Low Risk list
If phrase found in clause
Return LOW
Step 4: If no keyword found
Return LOW

5.5 Risk Scoring Module

The system calculates an overall document risk score using a weighted scoring mechanism:

- HIGH Risk → 3 points
- MEDIUM Risk → 2 points
- LOW Risk → 1 point

The score is computed using the formula:

Risk Score=∑(Clause Weight)Total Clauses×3×100Risk\ Score = \frac{\sum (Clause\ Weight)}{Total\ Clauses \times 3} \times 100Risk Score=Total Clauses×3∑(Clause Weight)×100

This produces a normalized percentage value ranging from 0 to 100.

5.6 Recommendation Engine

The recommendation engine analyzes detected risk patterns and generates actionable suggestions.

For example:

- Detection of indemnification clauses → Suggest review of liability scope
- Detection of termination clauses → Suggest negotiation of notice period

This module enhances decision-making by providing practical legal insights.

5.7 Web Interface Module

The system is implemented as a web application using the Flask framework. The interface provides:

- Drag-and-drop file upload
- Real-time analysis processing
- Risk score visualization (gauge)
- Clause-level breakdown
- Export/print report functionality

The front-end is designed using HTML, CSS, and JavaScript to ensure responsiveness and usability.

5.8 Implementation Advantages

- Lightweight and efficient (no heavy ML models)
- Fully explainable outputs
- Easy deployment and scalability
- Minimal hardware requirements
- User-friendly interface

6. Results And Discussion

The performance of the proposed ContractRisk Analyzer AI system was evaluated using multiple sample contract documents containing a mix of legal clauses with varying levels of risk. The system successfully identified and classified clauses into High, Medium, and Low risk categories based on predefined keyword patterns.

The results demonstrate that the system is capable of performing rapid and consistent analysis of contract documents while maintaining transparency in decision-making.

Table 1: Risk Classification Summary

Risk Level	Number of Clauses	Weight Assigned	Contribution
High	12	3	36
Medium	18	2	36
Low	20	1	20
Total	**50**	—	**92**

Overall Risk Score Calculation

Risk Score=9250×3×100=61.33%

Risk\ Score = \frac{92}{50 \times 3} \times 100 = 61.33\%Risk Score=50×392×100=61.33%

6.1 Discussion

The obtained results indicate that the system effectively identifies risk-prone clauses within legal documents. High-risk clauses such as indemnification and termination conditions were accurately detected and flagged, enabling users to focus on critical areas requiring attention.

The system provides several advantages:

- **Speed**: Processes entire contracts within seconds
- **Consistency**: Eliminates human errors and fatigue
- **Explainability**: Clearly highlights triggering keywords
- **Accessibility**: Requires minimal technical setup

However, certain limitations were observed:

- The system relies on keyword matching and may miss semantically complex risks
- It may generate false positives in cases involving negation
- Performance depends on the completeness of the risk dictionary

Despite these limitations, the system proves to be highly effective as a first-level contract screening tool, significantly reducing manual workload.

7. CONCLUSION

This paper presented ContractRisk Analyzer AI, a rule-based Natural Language Processing system designed to automate legal contract risk assessment. The system successfully integrates document processing, clause segmentation, risk classification, and recommendation generation into a unified framework.

The proposed solution addresses key challenges in traditional contract review by providing a fast, transparent, and cost-effective alternative. Unlike complex machine learning models, the system ensures complete explainability, making it suitable for real-world legal applications where interpretability is essential.

The results demonstrate that the system can significantly reduce review time while maintaining analytical accuracy. It enables users to quickly identify critical clauses and make informed decisions, thereby improving efficiency in legal and corporate environments.

Future enhancements may include:

- Integration of machine learning models for semantic analysis
- Expansion of the legal risk knowledge base
- Support for multilingual contract analysis
- Deployment as a cloud-based SaaS platform

Overall, the system represents a practical step towards intelligent legal automation and highlights the potential of explainable AI in the LegalTech domain.

Acknowledgement

The authors express their sincere gratitude to the faculty and management of Viswam Engineering College for their continuous support and encouragement in completing this research work.

8. REFERENCES

[1] Devlin J., Chang M.W., Lee K., Toutanova K., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019.
[2] Chalkidis I., Androutsopoulos I., Michos A., "Legal-BERT: The Muppets Straight Out of Law School," EMNLP, 2020.
[3] Hendrycks D., Burns C., Chen A., Ball S., "CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review," NeurIPS, 2021.
[4] Ashley K.D., "Artificial Intelligence and Legal Analytics," Cambridge University Press, 2017.
[5] Surden H., "Machine Learning and Law," Washington Law Review, 2014.

[6] Katz D.M., Bommarito M.J., Blackman J., "A General Approach for Predicting the Behavior of the Supreme Court," PLOS ONE, 2017.
[7] Dale R., "Law and Word Order: NLP Applications in Legal Text," Journal of AI and Law, 2019.
[8] Bird S., Klein E., Loper E., "Natural Language Processing with Python," O'Reilly Media, 2009.
[9] PyMuPDF Documentation, "PDF Processing with Python," 2023.
[10] Flask Documentation, "Web Development with Flask Framework," 2023.