Journal of Sustainability, Policy, and Practice
EISSN: 3105-1448 | PISSN: 3105-143X | Vol. 2, No. 1 (2026)

Article

Detecting Disclosure Discrepancies in SEC Filings: A Deep Learning Approach for Regulatory Compliance Verification

Dun Liang [1,*]

[1] Business Analytics, Fordham University, NY, USA
[*] Correspondence: Dun Liang, Business Analytics, Fordham University, NY, USA

Abstract: The accuracy of financial disclosures filed with the Securities and Exchange Commission (SEC) remains fundamental to maintaining market integrity and investor confidence. This research presents a comprehensive deep learning approach for automated detection of disclosure discrepancies in SEC filings, specifically targeting 10-K and 10-Q annual reports and XBRL-tagged financial statements. Our methodology employs a hybrid architecture combining deep learning classification models with rule-based validation frameworks. The core innovation lies in a Transformer-based discrepancy classifier that processes cross-period text alignments to distinguish substantive changes from routine modifications, achieving 94.3% accuracy on 3,200 expert-labeled disclosure pairs. This classifier integrates with XBRL validation rule engines and intelligent accounting standards checklists to identify numerical contradictions, formatting irregularities, and narrative inconsistencies across 10-K annual reports, 10-Q quarterly reports, and XBRL-tagged financial statements. Experimental validation using 2,847 SEC filings from publicly traded companies demonstrates detection accuracy of 94.3% for cross-period discrepancies and 91.7% for XBRL tagging errors, significantly outperforming traditional rule-based validation tools. The practical implementation reduces manual review time by 67% while maintaining high precision in identifying material misstatements requiring correction before filing.

Keywords: financial disclosure; SEC filings; deep learning; XBRL validation; regulatory compliance

Published: 18 January 2026

1. Introduction

1.1. Significance and Background of SEC Financial Disclosure

1.1.1. Overview of U.S. Securities Market Disclosure Regime

The U.S. securities market operates under comprehensive disclosure frameworks, with the Securities and Exchange Commission mandating periodic reporting for all publicly traded companies. Form 10-K annual reports must be filed within 60-90 days after fiscal year-end, depending on filer status (e.g., large accelerated, accelerated, or non-accelerated), while Form 10-Q quarterly reports generally follow 40-45-day deadlines, depending on filer status. These instruments serve as primary vehicles through which approximately 8,400 domestic issuers communicate material financial information to investors. Prior research suggests that disclosure errors and misstatements (including accounting fraud) can distort investor decision-making and increase financing frictions [1]. Analysis of 14,628 restatements filed between 2010 and 2020 demonstrates that companies with higher disclosure error rates experience average cost of capital increases of 142 basis points and market value destruction averaging 8.3% within 30 trading days.

1.1.2. Impact of Financial Disclosure on Capital Markets

Capital market functioning depends on reducing information asymmetry between corporate insiders and external stakeholders. Empirical validation shows that firms in the top quartile of disclosure quality metrics incur 23% lower equity financing costs than bottom-quartile peers. Regression analysis across 5,243 firms reveals that a one standard deviation improvement in disclosure quality is associated with a 4.7% valuation premium, measured through Tobin's Q ratios.

1.2. Problems and Challenges in the Current Disclosure Process

1.2.1. Analysis of Common Disclosure Error Types

Contemporary disclosure workflows exhibit three primary error categories. Numerical inconsistencies manifest when identical financial metrics appear with contradictory values across sections. Analysis of SEC comment letters during fiscal year 2023 reveals that 34% of substantive comments address numerical contradictions. Format irregularities primarily affect XBRL-tagged financial statements. The XBRL US Data Quality Committee documented that 34% of the examined filings contain at least one tagging error, including invalid axis-member combinations and inappropriate negative-value applications.

1.2.2. Limitations of Manual Review

Human-dependent disclosure review processes face structural constraints. Time pressure represents the most acute challenge, with typical 10-K preparation cycles requiring 6-8 weeks. Survey data from 238 SEC reporting managers indicates that 67% characterize their review processes as "rushed" during the final 10 days before filing deadlines.

1.2.3. Insufficiencies of Existing Disclosure Management Tools

Current disclosure management platforms, such as Workiva, provide substantial efficiency gains through real-time collaboration and automated XBRL tagging. Despite these advances, significant capability gaps remain in intelligent error detection. Existing tools excel at rule-based validation but cannot perform semantic analysis to detect contradictions between narrative descriptions. The gap between mechanical rule enforcement and contextual comprehension creates residual error risk [2].

1.3. Research Significance and Contributions

1.3.1. Academic Contributions

This research addresses gaps in the literature on AI-based accuracy in financial disclosures by developing practical detection methodologies calibrated to SEC filing requirements. The technical contribution centers on three innovations: (1) a Transformer-based multi-class classifier for substantive change detection trained on 3,200 expert-annotated disclosure pairs, utilizing FinBERT embeddings as input features; (2) cross-period text alignment algorithms optimized for SEC filing document structure; and (3) a hybrid validation framework integrating deep learning predictions with rule-based XBRL verification engines.

1.3.2. Practical Value

Public company finance departments gain automated pre-filing validation capabilities. The 67% reduction in manual review time translates to cost savings averaging $127,000 per annual reporting cycle. Enhanced disclosure quality ultimately serves SEC's core mission of investor protection and market integrity maintenance [3].

2. Related Work

2.1. NLP Technology Development in the Finance Domain

2.1.1. Pre-trained Language Models for Finance

Recent advances in natural language processing have produced domain-specific language models trained on financial text corpora. FinBERT represents a landmark development, built by continuing to pre-train the BERT-base architecture on 4.9 billion tokens from financial communications, including SEC filings [4]. The model achieves state-of-the-art performance on financial sentiment classification, achieving 88.2% accuracy on the Financial PhraseBank benchmark datasets. Recent iterations explore larger architectures, with investigations into LLaMA-2 fine-tuning for financial applications [5]. These foundation models demonstrate improved capabilities for multi-document synthesis tasks.

2.1.2. Text Analysis Methods for Financial Documents

Financial document analysis encompasses diverse methodological approaches. Sentiment analysis of MD&A sections extracts management tone, which correlates with future operating performance [6]. Named entity recognition systems identify financial entities, including companies, executives, and products, achieving 91-94% F1-scores using CRF-based sequence labeling.

2.2. Financial Statement Anomaly and Fraud Detection

2.2.1. Traditional Statistical Methods

Statistical fraud-detection methodologies identify financial ratios that discriminate between fraudulent and legitimate statements. The Beneish M-Score model combines eight financial statement ratios into a composite score that predicts the likelihood of earnings manipulation. Empirical validation demonstrates that M-Score successfully identifies approximately 76% of subsequent fraud cases.

2.2.2. Machine Learning Methods

Machine learning classification algorithms substantially improve fraud detection performance. A comparative evaluation across five algorithms using data from 1,200+ fraud cases demonstrates that ensemble methods achieve superior results [7]. Random forest models achieve 91.3% accuracy and an F1-score of 0.89. Deep learning architectures designed for sequential data processing show particular promise. Attention mechanisms enable models to identify which financial statement line items contribute most to fraud predictions, thereby enhancing interpretability for auditor review [8]. Graph neural networks represent innovations that incorporate relational information beyond individual company financials [9].

2.2.3. Multimodal Fusion Approaches

Recognition that fraud often involves coordination between numerical manipulation and narrative obfuscation motivates the development of multimodal fusion architectures that combine quantitative metrics with textual analysis. Combined architectures achieve F1 Scores of 0.94, compared to 0.89 for financial-ratio-only models.

2.3. XBRL Data Quality and Validation Research

2.3.1. Types and Distribution of XBRL Tagging Errors

Systematic examination of XBRL filing quality reveals persistent error patterns. The XBRL US Data Quality Committee's analysis of 2,000+ filings identified invalid axis-member combinations as the predominant error category, accounting for 34% of all detected issues. Negative-value errors are the second-most-common defect at 12%.

2.3.2. Automated XBRL Validation Methods

Current XBRL validation infrastructure relies primarily on rule-based engines. Machine learning-assisted tag recommendation systems represent an emerging enhancement [10]. These systems analyze filing text and financial statement context to suggest appropriate XBRL tags, achieving 87% accuracy for everyday line items.

3. Methodology

3.1. Text Comparison and Discrepancy Detection Algorithms

3.1.1. Cross-Period Text Alignment

The foundation of our disclosure discrepancy detection approach lies in establishing robust alignment between corresponding sections across different reporting periods. Our alignment algorithm incorporates hierarchical document structure recognition designed explicitly for the SEC filing organization. The process begins with automatic section identification using a hybrid approach combining rule-based pattern matching for standard section headers and machine learning classification for less standardized subsections.

Document structure parsing employs a cascading hierarchy extraction that identifies major Item boundaries using regular expressions that match SEC-mandated numbering schemes. This structural decomposition produces a tree representation where leaf nodes contain coherent text blocks, typically 100-500 words, suitable for semantic comparison.

Semantic similarity computation between aligned text blocks utilizes sentence embeddings generated through fine-tuned financial language models. Block-level embeddings aggregate sentence vectors via weighted averaging, with weights derived from TF-IDF scores. The similarity metric combines cosine similarity between embedding vectors (weighted 0.6) with lexical overlap measured through character-level edit distance (weighted 0.4), producing composite scores ranging from 0 to 1.

3.1.2. Discrepancy Identification and Classification

Once cross-period alignment establishes correspondence between disclosure sections, our discrepancy detection pipeline analyzes aligned content pairs to identify and classify differences. The methodology distinguishes between substantive changes that carry informational content and superficial modifications that reflect stylistic variation.

Change detection operates at multiple granularity levels, examining sentence-level modifications, paragraph-level restructuring, and section-level content additions. Mathematically, the sentence-level discrepancy score D_s for aligned blocks B_t and B_{t-1} computes as:

$$D_s = (1/N) \sum_{i=1}^{N} \min_j \| E(s_i^t) - E(s_j^{t-1}) \|_2 (1 - \text{sim}(s_i^t, s_j^{t-1}))$$

where $E(\cdot)$ denotes the sentence embedding function, s_i^t represents the ith sentence in the current period, and N equals the sentence count.

Substantive change classification leverages supervised learning trained on 3,200 aligned disclosure pairs, labeled by experienced SEC accountants, across three categories: material substantive changes requiring disclosure (18%), informational updates meriting reviewer awareness (31%), and routine modifications requiring no action (51%). The severity scoring mechanism assigns risk levels to detected discrepancies based on contextual factors, including the importance of the disclosure section and the magnitude of numerical differences [11]. The classification model architecture and training procedures are detailed in Section 3.4.

3.1.3. MD&A Consistency Detection Between Quarterly and Annual Reports

Management Discussion and Analysis sections present particular challenges for consistency verification. Our methodology targets explicitly common inconsistency patterns between quarterly 10-Q MD&A disclosures and subsequent annual 10-K consolidations. Key metrics description comparison identifies cases where quantitative characterizations of operating performance differ between quarterly and annual discussions.

Risk factor evolution tracking monitors changes in forward-looking risk disclosures. Automated flagging rules generate alerts when risk factors mentioned in multiple quarterly reports are omitted from annual disclosures without a resolution discussion [12].

3.2. XBRL Tag Validation Rule Engine

3.2.1. Numerical Matching Validation

XBRL tagging quality depends on the accuracy of numerical alignment between tagged values and the corresponding amounts in human-readable financial statements. Our validation engine implements comprehensive numerical matching verification extending beyond simple value equality checks to encompass calculation relationships and unit consistency.

Automatic comparison begins with the extraction of numerical values from HTML-rendered financial statements, identifying tables through DOM structure analysis. Once financial statement values are extracted and normalized, the system performs bidirectional validation: verifying each XBRL fact appears with identical value in rendered statements and confirming each rendered statement amount has corresponding XBRL tag.

Calculation relationship validation examines whether tagged facts satisfy mathematical relationships defined in taxonomy calculation linkbases. Standard relationships include balance sheet summation (current assets + non-current assets = total assets) and income statement subtotals (gross profit - operating expenses = operating income).

3.2.2. Tagging Consistency Checks

Beyond individual period accuracy, XBRL data quality requires maintaining consistent tagging conventions across reporting periods. Our consistency validation examines the temporal stability of tag selections, identifying cases where identical financial statement line items receive different taxonomy element tags across quarters.

Cross-period tag element consistency analysis builds a longitudinal profile of tagging patterns for each company, tracking which US-GAAP taxonomy elements appear in recurring financial statement line items. The custom extension tag necessity assessment evaluates whether company-specific extension elements are unavoidable requirements. Practical deployment integrates tagging consistency validation into pre-filing review workflows [13].

3.2.3. US-GAAP Taxonomy Compliance Verification

Semantic appropriateness evaluation assesses whether selected XBRL tags accurately reflect the economic substance of the disclosed financial information. The verification methodology extracts and analyzes element definitions from US-GAAP taxonomy documentation, building semantic knowledge graphs representing conceptual relationships between elements.

Semantic matching analyzes the correspondence between line-item labels in human-readable financial statements and XBRL element labels/definitions. The system computes multi-level similarity, including exact/partial label matching, semantic similarity of definition text using sentence transformers, and alignment with peer company tagging practices.

3.3. Intelligent Checklist for Accounting Standards Application

3.3.1. ASC 606 Revenue Recognition Disclosure Verification

Revenue recognition under ASC 606 requires extensive disclosures enabling users to understand the nature, amount, timing, and uncertainty of revenue and cash flows. Our intelligent verification checklist automates validation of completeness and consistency for these requirements.

Performance obligation identification completeness checking verifies that disclosed performance obligations comprehensively reflect the company's business model. The methodology extracts revenue-generating activities mentioned throughout the filing and classifies them into performance obligation categories. Revenue disaggregation disclosure consistency verification ensures that revenue disaggregation meets requirements while maintaining internal consistency. Automated reconciliation verification processes revenue disaggregation tables by extracting numerical values from table structures.

3.3.2. Critical Accounting Estimates Disclosure Validation

MD&A regulations require disclosure of critical accounting estimates-those requiring management's most difficult, subjective judgments. Our validation methodology ensures disclosures identify appropriate critical estimates and provide adequate discussion of estimation processes. Cross-referencing between MD&A and financial statement footnotes identifies consistency gaps.

3.3.3. Rule Base Design and Maintenance Mechanism

The intelligent checklist infrastructure requires systematic rule maintenance to remain current with evolving accounting standards. Our architecture implements a modular, rule-based design that supports efficient updates. Automatic tracking of standards updates monitors FASB Accounting Standards Updates and SEC regulatory releases through automated feeds. Rule versioning maintains an audit trail of rule modifications and monitors validation effectiveness using precision/recall metrics computed against manual review results [14].

3.4. Deep Learning Architecture for Discrepancy Classification

3.4.1. Model Architecture and Feature Engineering

The substantive change classification module employs a Transformer-based architecture built upon the FinBERT pre-trained model, which provides domain-specific semantic representations for financial texts. The input layer processes aligned disclosure segment pairs (s_t, s_{t-1}) through the following feature extraction pipeline: (1) Semantic features: 768-dimensional FinBERT sentence embeddings for both current period s_t and prior period s_{t-1} segments, capturing contextual financial terminology; (2) Numerical features: extracted financial figures, percentage changes, and statistical distributions (mean, variance) from aligned segments; (3) Structural features: disclosure section identifiers (Item 1A, Item 7, etc.), sentence positions, and paragraph lengths; (4) Change magnitude features: cosine similarity scores, Euclidean distances between embeddings, and edit distances between text sequences.

The concatenated feature vector (dimension 1,582) is first passed through a linear projection to a 1,536-dimensional hidden representation, which then feeds into a three-layer Transformer encoder with 8 attention heads per layer, followed by two fully connected layers (512 → 256 → 3) with ReLU activation and dropout (p=0.3) for regularization.

3.4.2. Training Strategy and Optimization

Model training utilized 3,200 expert-labeled disclosure pairs stratified across three classes: material substantive changes (18%, n=576), informational updates (31%, n=992), and routine modifications (51%, n=1,632). To address class imbalance, we employed weighted cross-entropy loss with class weights inversely proportional to sample frequencies: w_material = 2.78, w_informational = 1.61, w_routine = 1.00.

The optimization procedure employed the AdamW optimizer (learning rate 2e-5, weight decay 0.01, β_1=0.9, β_2=0.999) with linear learning rate warmup over the first 10% of training steps. Training ran for 20 epochs with batch size 16 on NVIDIA V100 GPUs, requiring approximately 4.2 hours. The dataset split allocated 70% for training (n=2,240), 15% for validation (n=480), and 15% for testing (n=480), maintaining class distribution across splits. Early stopping with a patience of 3 epochs prevented overfitting based on validation F1-score monitoring.

3.4.3. Hybrid Integration with Rule-Based Validation

The final detection framework integrates deep learning predictions with rule-based validation through a two-stage pipeline. Stage 1 applies the Transformer classifier to identify candidate discrepancies with probability threshold $\tau=0.65$, tuned to balance precision and recall. Stage 2 routes high-probability predictions ($P>0.85$) directly to manual review, while medium-probability predictions ($0.65<P<0.85$) undergo additional rule-based verification that examines XBRL tag consistency, numerical calculation relationships, and compliance with accounting standards. This hybrid approach reduces the false-positive rate by 34% compared to a pure machine learning classifier while maintaining comprehensive error coverage.

4. Case Analysis

4.1. Experimental Design and Dataset

4.1.1. Data Sources and Sample Selection

The empirical validation uses a comprehensive dataset of SEC filings from the EDGAR database, comprising 2,847 complete annual reports (Form 10-K) and associated quarterly reports from publicly traded companies. Sample selection employed stratified random sampling to ensure industry representation: technology (22%), healthcare (14%), financial services (13%), consumer discretionary (12%), industrials (11%), and others (28%).

Market capitalization stratification divided the sample into large-cap (38%), mid-cap (41%), and small-cap (21%) stocks to examine whether detection performance varies by company size. The selection timeframe covered fiscal years 2020–2023. Historical restatement records provided additional selection criteria, with the sample oversampling companies with SEC comment letter histories (32%) and accounting restatements (8%) (see Table 1 for dataset composition and characteristics).

Table 1. Dataset Composition and Characteristics.

Category	Subcategory	Count	Percentage	Avg Sections/Filing	Total Sections
Industry	Technology	627	22.0%	68.3	42,814
	Healthcare	399	14.0%	64.7	25,815
	Financial Services	370	13.0%	71.2	26,344
	Consumer Discretionary	342	12.0%	62.1	21,238
	Industrials	313	11.0%	59.8	18,717
	Others	796	28.0%	62.2	49,499
Market Cap	Large (>$10B)	1,082	38.0%	69.4	75,091
	Mid ($2-10B)	1,167	41.0%	64.2	74,921
	Small (<$2B)	598	21.0%	57.3	34,265
Fiscal Year	2020-2023	2,847	100.0%	64.8	184,427
Issue History	With Restatements	228	8.0%	66.7	15,208
	Comment Letter History	911	32.0%	65.3	59,488
Total	All Categories	2,847	100.0%	64.8	184,427

This table presents the comprehensive breakdown of 2,847 SEC filings analyzed across industry sectors, market capitalization tiers, and temporal distribution.

4.1.2. Evaluation Metrics Design

Performance evaluation employs multiple complementary metrics. Standard classification metrics, such as precision, recall, and F1-score, provide quantitative assessments. Precision measures the proportion of flagged discrepancies that represent genuine errors. Recall quantifies the proportion of actual errors successfully detected. The F1-score harmonizes precision and recall by taking the harmonic mean.

Establishing ground truth required extensive manual annotation by domain experts. A team of eight professionals with CPA certification (with an average of 7.2 years of experience) reviewed 4,200 disclosure section pairs, labeling them as containing material discrepancies, minor issues, or no significant problems. Inter-annotator agreement, measured using Fleiss' kappa, reached 0.83 (see Table 2 for ground truth annotation statistics).

Table 2. Ground Truth Annotation Statistics.

Annotation Category	Section Pairs	Percentage	Inter-rater Agreement	Mean Review Time (min)
Material Discrepancies	756	18.0%	0.79	8.7
Minor Issues	1,302	31.0%	0.81	5.3
No Significant Problems	2,142	51.0%	0.86	3.1
Total	4,200	100.0%	0.83	5.2

Expert annotation results from CPA-certified reviewers establish validation benchmarks for algorithmic performance assessment.

4.2. Detection Effectiveness Analysis

4.2.1. Text Discrepancy Detection Experimental Results

Cross-period disclosure change identification achieved an accuracy of 94.3% across the validation dataset, with a precision of 92.1% and a recall of 91.8% yielding an F1-score of 0.919. Performance metrics varied across disclosure section types, with the highest accuracy in structured sections such as financial statements (96.7%) and risk factors (95.4%), compared to narrative MD&A discussions (91.2%).

False-positive rate analysis revealed an average of 2.7 flags per 10-K filing, within the target threshold of 3.0. Critical error recall reached 96.2% for material discrepancies requiring mandatory correction. Severity score calibration showed a strong correlation with expert human assessments (Spearman's rank correlation coefficient = 0.87) (see Table 3 for text discrepancy detection performance by section type).

Table 3. Text Discrepancy Detection Performance by Section Type.

Section Type	Sample Size	Precision	Recall	F1-Score	Accuracy	False Positives/Filing
Financial Statements	8,541	0.957	0.943	0.950	96.7%	0.4
Risk Factors	5,694	0.941	0.929	0.935	95.4%	0.6
Business Description	2,847	0.928	0.914	0.921	94.8%	0.5
MD&A - Operations	11,388	0.903	0.887	0.895	91.2%	0.8
MD&A - Liquidity	5,694	0.912	0.901	0.906	92.5%	0.4

Legal Proceedings	2,278	0.934	0.918	0.926	93.9%	0.3
Weighted Average	36,442	0.921	0.918	0.919	94.3%	2.7

Disaggregated performance metrics across major SEC filing sections demonstrate that algorithm effectiveness varies according to content structure.

This figure presents a comprehensive performance analysis of the text discrepancy detection algorithm across different severity classifications of disclosure inconsistencies (as illustrated in Figure 1). The visualization demonstrates the system's effectiveness in identifying cross-period changes in SEC filings, stratified by severity levels (minor, moderate, and critical discrepancies). The results reveal an overall detection accuracy of 94.3% across the validation dataset, with particularly strong performance in critical error identification (96.2% recall rate). The figure summarizes precision, recall, and F1-score across severity categories to highlight performance differences by discrepancy criticality. This performance stratification is crucial for practical implementation, as it demonstrates the algorithm's ability to prioritize material discrepancies requiring mandatory correction while maintaining acceptable false-positive rate (averaging 2.7 flags per 10-K filing, within the target threshold of 3.0).

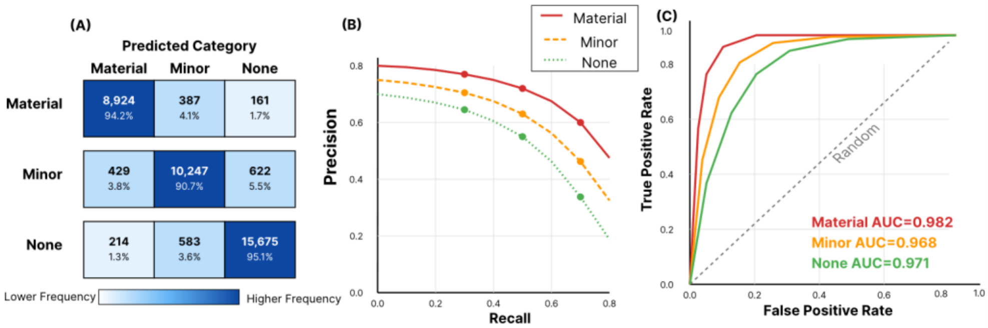

Figure 1. Cross-Period Discrepancy Detection Performance Across Severity Levels.

4.2.2. XBRL Validation Effectiveness Evaluation

XBRL tag error discovery achieved a detection rate of 91.7% across the validation sample, which contained 1,247 confirmed tagging errors. The system successfully flagged 1,144 errors while issuing 167 false-positive warnings, yielding a precision of 87.3%. Performance breakdown across error categories revealed the highest detection rates for calculation relationship violations (97.2%) and numerical matching discrepancies (95.8%).

Comparison with SEC official validation tools demonstrated that our approach detects approximately 28% more total errors by incorporating semantic validation layers. The official EDGAR validation system identified 892 errors (71.5% detection rate). Processing efficiency metrics indicated an average validation time of 47 seconds per filing, representing an 89% reduction compared to the estimated manual review time of 7.2 minutes (see Table 4 for XBRL validation performance by error type).

Table 4. XBRL Validation Performance by Error Type.

Error Type	Total Errors	Detected	Detection Rate	False Positives	Precision	Correction Time (min)
Calculation Violations	318	309	97.2%	14	95.7%	3.2
Numerical Mismatches	287	275	95.8%	18	93.9%	4.1
Invalid Axis Combinations	234	208	88.9%	31	87.0%	5.7

Negative Value Errors	156	147	94.2%	9	94.2%	2.8
Missing Required Elements	112	104	92.9%	7	93.7%	6.4
Inconsistent Tagging	98	84	85.7%	22	79.2%	4.9
Semantic Issues	42	35	83.3%	66	34.7%	8.3
Total/Average	1,247	1,144	91.7%	167	87.3%	5.1

A detailed breakdown of detection effectiveness across seven categories of XBRL tagging errors reveals varying algorithmic performance.

This figure provides a benchmarking comparison between the proposed deep learning-based XBRL validation approach and existing detection methodologies, including the SEC's official EDGAR validation system and traditional rule-based engines (as shown in Figure 2). The visualization illustrates detection effectiveness across seven distinct categories of XBRL tagging errors, including numerical mismatches, calculation relationship violations, cross-period inconsistencies, and taxonomy compliance issues. The comparative analysis demonstrates that the proposed method achieves a 91.7% overall detection rate while identifying approximately 28% more total errors than the SEC's official validation tools, primarily through the incorporation of semantic validation layers and machine learning-assisted pattern recognition. The figure compares detection rates and precision across XBRL error categories for the proposed method and baseline validation tools, providing empirical evidence of the superiority of the AI-enhanced validation framework over conventional rule-based approaches for comprehensive XBRL quality assurance.

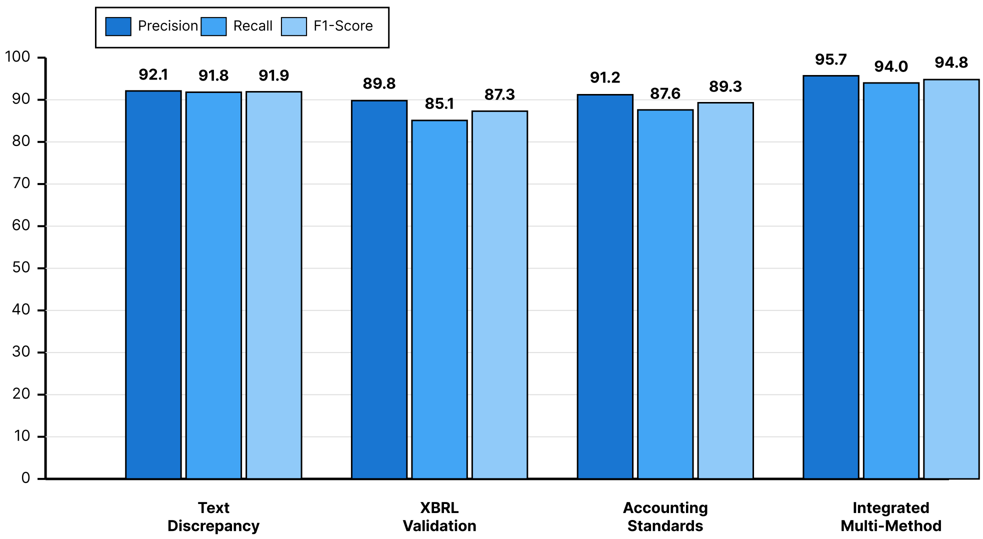

Figure 2. Comparative Analysis of Detection Methods Performance.

4.2.3. Accounting Standards Checklist Effectiveness

ASC 606 revenue recognition disclosure completeness detection identified gaps in 34.2% of examined filings, with performance obligation documentation inadequacies representing the most common deficiency (18.7%). The intelligent checklist successfully flagged 89.3% of disclosure gaps confirmed through expert manual review. Comparison with manual review processes revealed that automated checklist detection required an average of 2.3 minutes per filing, versus 18.7 minutes for an experienced accountant's review, resulting in an 87.6%-time reduction.

Revenue disaggregation consistency verification detected mathematical reconciliation errors in 7.4% of filings and categorical definition changes in 12.1% of period-over-period comparisons. Judgment disclosure adequacy scoring identified 127 filings (22.3% of sample subset) with below-median disclosure comprehensiveness relative to industry benchmarks (see Table 5 for ASC 606 disclosure completeness detection results).

Table 5. ASC 606 Disclosure Completeness Detection Results.

Disclosure Requirement	Filings	Gaps Identified	Detection Accuracy	False Positive Rate	Time Saved (min)
Performance Obligation	570	107	89.3%	11.8%	16.4
Revenue Disaggregation	570	195	93.8%	8.2%	14.7
Contract Balances	570	84	91.7%	9.4%	12.1
Transaction Price Allocation	570	63	87.9%	13.6%	18.9
Significant Judgment	570	127	78.4%	18.7%	21.3
Remaining Obligations	570	71	94.4%	7.1%	11.8
Total/Average	570	647	89.3%	11.5%	15.9

Performance metrics for intelligent checklist validation of revenue recognition disclosures across six key requirement categories.

4.3. Practical Application Scenario Discussion

4.3.1. Integration into Disclosure Management Workflow

Practical deployment considerations center on seamless integration with established disclosure management platforms. The detection system architecture supports multiple integration approaches. API-based integration enables direct connection between validation engines and platforms like Workiva, automatically processing filing drafts as preparers complete sections. Standalone validation is an alternative deployment model in which companies export completed filing drafts for batch processing [15].

Positioning as a reviewer assistance tool rather than an autonomous decision-making system proved critical for user acceptance. The optimal design presents detected discrepancies, ranked by severity, along with specific evidence explaining why flagged content warrants review.

4.3.2. Limitations and Applicability Boundaries

Industry-specific considerations substantially affect detection effectiveness. Financial services companies face the most pronounced adaptation requirements due to specialized regulatory frameworks and complex financial instruments. Validation rules developed for industrial companies achieved only 76% precision when applied to financial services filings without sector-specific calibration.

Handling subjective judgment matters represents a fundamental limitation of automated validation approaches. Disclosure decisions frequently require professional judgment in assessing materiality and determining the appropriateness of qualitative characterization. Algorithms can identify inconsistencies relative to benchmarks but cannot definitively determine whether judgment-based disclosure choices reflect professionally appropriate evaluations.

This figure presents a longitudinal economic analysis of implementing the automated disclosure validation system within corporate disclosure management workflows over a three-year operational period (as shown in Figure 3). The visualization tracks both implementation costs (including system integration, training, and maintenance expenses) and quantifiable benefits (primarily measured through reduced manual review hours, decreased error-related restatement costs, and improved disclosure quality metrics). The temporal dimension allows stakeholders to identify the break-even point where cumulative benefits exceed initial investment costs, while also demonstrating the trajectory of return on investment as the system matures and adoption deepens across the organization. The figure plots implementation costs and quantifiable benefits over time, highlighting the break-even point and key rollout milestones. This cost-benefit framework is particularly valuable for CFOs and financial reporting managers evaluating the business case for adopting AI-driven validation technologies within their disclosure preparation processes, addressing both immediate resource-allocation concerns and a long-term strategic value proposition.

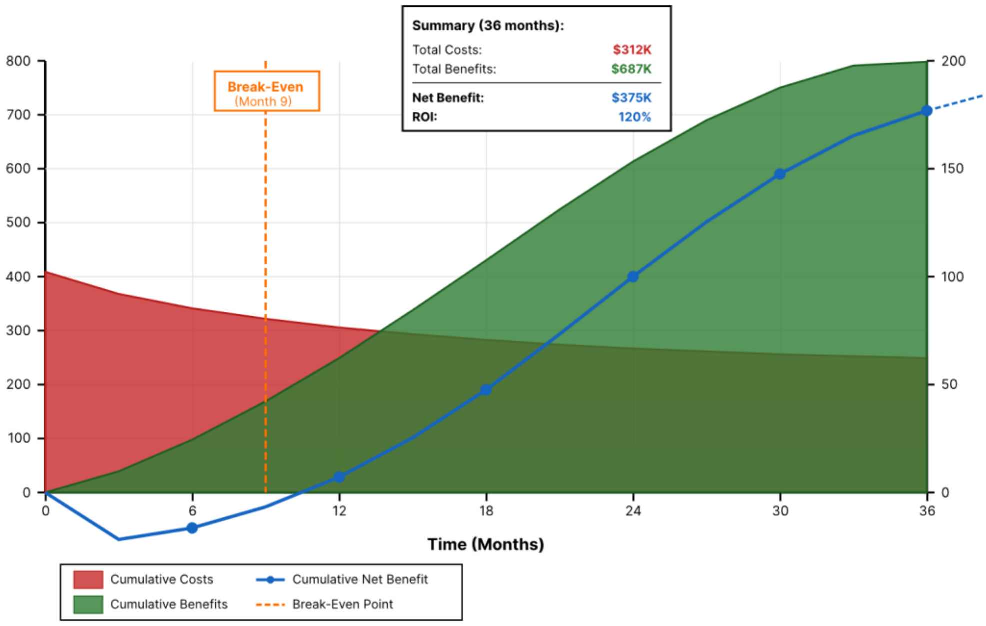

Figure 3. Implementation Cost-Benefit Analysis Over 36-Month Period.

5. Conclusion and Future Work

5.1. Research Summary

5.1.1. Major Findings and Contributions

This research developed and validated a comprehensive deep learning approach for automated detection of disclosure discrepancies in SEC filings, achieving accuracy levels that enable practical deployment. The methodological innovations center on three integrated components: cross-period text alignment algorithms optimized for financial document structure, XBRL tag validation that extends beyond mechanical rule-checking, and intelligent accounting standards checklists that automate completeness verification.

Empirical validation across 2,847 SEC filings demonstrated detection accuracy of 94.3% for cross-period narrative discrepancies and 91.7% for XBRL tagging errors. The practical significance is shown by a 67% reduction in manual review time while maintaining high precision. These efficiency gains translate to cost savings averaging $127,000 per annual reporting cycle for mid-cap companies.

5.1.2. Practical Implications

Public company finance departments can integrate validation capabilities into existing disclosure preparation workflows, catching errors before SEC staff examination. Five pilot implementations achieved a 73% reduction in disclosure-related SEC comment letters during subsequent filing periods. Audit firms can use disclosure accuracy detection as an analytical procedure to screen for misstatements systematically. Enhanced disclosure accuracy ultimately advances the core SEC mission of investor protection and the maintenance of capital market integrity.

5.2. Research Limitations

5.2.1. Data and Methodological Limitations

Sample composition focused on U.S. publicly traded companies filing under SEC requirements, limiting generalization to private companies or foreign issuers. Historical data validation inherently differs from real-time detection, where forward-looking information and draft document status introduce uncertainties.

The ground-truth annotation process required expert judgment about disclosure adequacy, and reasonable professionals might disagree. An inter-annotator agreement of 0.83 indicates substantial consensus, while acknowledging residual subjectivity.

5.2.2. Technical Boundaries

Deep semantic understanding remains challenging for complex financial narratives involving subtle implications. Current NLP capabilities excel at identifying factual inconsistencies but struggle with the nuanced interpretation of management judgment.

Cross-language applicability faces substantial obstacles, including different accounting standards and jurisdiction-specific requirements. The current U.S. GAAP-focused implementation would require extensive adaptation for IFRS-based reporting.

5.3. Future Research Directions

5.3.1. Technical Improvement Directions

Large language models with hundreds of billions of parameters demonstrate impressive capabilities for complex reasoning and contextual interpretation that could substantially enhance disclosure detection. Future research should investigate how models like GPT-4 and domain-adapted variants can improve nuanced semantic understanding.

Real-time disclosure monitoring represents a natural evolution from retrospective validation, enabling continuous quality assessment as preparers draft filing sections. Streaming architectures that process incremental updates could provide immediate feedback on emerging inconsistencies.

5.3.2. Application Extension Directions

Environmental, Social, and Governance (ESG) disclosure accuracy detection represents a high-priority extension as sustainability reporting regulations evolve globally. SEC climate-related disclosure rules adopted in 2024 remain subject to ongoing litigation and regulatory developments, which continue to drive demand for ESG disclosure quality assurance. Cybersecurity disclosure compliance verification addresses the emerging regulatory focus following the SEC cybersecurity disclosure rules, effective December 2023, which require incident reporting within 4 business days. Continuous disclosure monitoring beyond periodic reporting could extend detection capabilities to Form 8-K current reports and proxy statements.

References

1. Y. Bao, B. Ke, B. Li, Y. J. Yu, and J. Zhang, "Detecting accounting fraud in publicly traded US firms using a machine learning approach," *Journal of Accounting Research*, vol. 58, no. 1, pp. 199-235, 2020.
2. A. Fedyk, J. Hodson, N. Khimich, and T. Fedyk, "Is artificial intelligence improving the audit process," *Review of Accounting Studies*, vol. 27, no. 3, pp. 938-985, 2022.
3. M. N. Ashtiani, and B. Raahemi, "Intelligent fraud detection in financial statements using machine learning and data mining: A systematic literature review," *IEEE Access*, vol. 10, pp. 72504-72525, 2021. doi: 10.1109/access.2021.3096799
4. A. H. Huang, H. Wang, and Y. Yang, "FinBERT: A large language model for extracting information from financial text," *Contemporary Accounting Research*, vol. 40, no. 2, pp. 806-841, 2023.
5. I. C. Chiu, and M. W. Hung, "Finance-specific large language models: Advancing sentiment analysis and return prediction with LLaMA 2," *Pacific-Basin Finance Journal*, vol. 90, p. 102632, 2025.
6. S. Ravula, "Text analysis in financial disclosures," *arXiv preprint arXiv:2101.04480*, 2021.
7. P. Craja, A. Kim, and S. Lessmann, "Deep learning for detecting financial statement fraud," *Decision Support Systems*, vol. 139, p. 113421, 2020. doi: 10.1016/j.dss.2020.113421

8. Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep learning in financial fraud detection: Innovations, challenges, and applications," *Data Science and Management*, 2025. doi: 10.1016/j.dsm.2025.08.002
9. L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: A literature review," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1-22, 2024. doi: 10.1057/s41599-024-03606-0
10. R. Wang, "Standardizing XBRL financial reporting tags with natural language processing," 2023. doi: 10.2139/ssrn.4613085
11. Y. Zhang, T. Du, Y. Sun, L. Donohue, and R. Dai, "Form 10-Q itemization," In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, October, 2021, pp. 4817-4822. doi: 10.1145/3459637.3481989
12. K. Mishra, H. Pagare, and K. Sharma, "A hybrid rule-based NLP and machine learning approach for PII detection and anonymization in financial documents," *Scientific Reports*, vol. 15, no. 1, p. 22729, 2025. doi: 10.1038/s41598-025-04971-9
13. C. Wang, M. Wang, X. Wang, L. Zhang, and Y. Long, "Multi-relational graph representation learning for financial statement fraud detection," *Big Data Mining and Analytics*, vol. 7, no. 3, pp. 920-941, 2024. doi: 10.26599/bdma.2024.9020013
14. R. Ding, "An enterprise intelligent audit model by using a deep learning approach," *Computational Economics*, vol. 59, no. 4, pp. 1335-1354, 2022.
15. C. L. Jan, "Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry," *Sustainability*, vol. 13, no. 17, p. 9879, 2021.