10.48047/jocaaa.2024.33.05.35

# A Large Language Model Framework for Intelligent Insurance Claim Automation and Fraud Detection

**Santosh Kumar**
Highlands Ranch, Colorado, USA, 80130
santosh.iimc07@gmail.com

## Abstract

The insurance industry, a cornerstone of global financial stability, is undergoing a profound transformation driven by artificial intelligence. Traditional claim processing remains a labor-intensive, time-consuming, and error-prone endeavor, often struggling to effectively identify sophisticated fraudulent activities. This paper proposes a novel, integrated framework that leverages the advanced capabilities of Large Language Models (LLMs) to automate and secure the end-to-end insurance claim lifecycle. Our framework systematically employs LLMs for critical tasks, including the automated ingestion and comprehension of multi-format claim documents (e.g., reports, invoices, medical records), intelligent triage and routing of claims based on complexity and risk, and the generation of human-readable summaries for adjuster review. Crucially, the model is fine-tuned to perform deep semantic analysis and anomaly detection across structured and unstructured data, identifying subtle indicators of fraudulent patterns that elude conventional rule-based systems. A prototype implementation demonstrates a significant reduction in processing times and operational costs, while simultaneously enhancing fraud detection accuracy. The findings posit that the proposed LLM-centric framework represents a paradigm shift, offering a scalable, efficient, and robust solution for the future of intelligent insurance operations.

**Keywords:** Large Language Models, Insurance Claim Automation, Fraud Detection, Natural Language Processing, Anomaly Detection, Intelligent Document Processing.

## 1. Introduction

### 1.1 Overview

The global insurance sector, a pivotal component of the world's economic infrastructure, is fundamentally predicated on the efficient and accurate assessment of risk and the subsequent management of claims. However, the core process of insurance claim adjudication remains mired in operational inefficiencies and vulnerabilities. Conventional workflows are characterized by extensive manual intervention, requiring human adjusters to parse complex, multi-modal information from First Information Reports (FIRs), medical certificates, mechanic assessments, and financial invoices. This labor-intensive paradigm inevitably leads to protracted processing times, elevated operational expenditures, and a non-trivial margin for human error. Concurrently, the industry faces a persistent and evolving threat from fraudulent claims, which, according to industry analyses, account for a significant percentage of all claims payouts, ultimately inflating premiums for honest policyholders and eroding profitability for insurers. Traditional rule-based systems and human-expertise-dependent models for fraud detection are increasingly proving inadequate against sophisticated, collusive, and adaptively engineered fraudulent schemes that leave minimal deterministic footprints.

10.48047/jocaaa.2024.33.05.35

The recent paradigm shift in artificial intelligence, catalyzed by the advent of Large Language Models (LLMs), presents a transformative opportunity to address these long-standing challenges. Models such as BERT [1], GPT-series [2], and T5 [3], built upon the Transformer architecture [4], have demonstrated unprecedented capabilities in natural language understanding (NLU), natural language generation (NLG), and complex reasoning across diverse domains. Their proficiency in comprehending semantic context, syntactic structures, and relational logic within unstructured text data positions them as ideal candidates for automating and securing the intellectual workflows inherent in insurance claim processing. This paper posits that a holistic framework built upon these models can move beyond mere automation to create an intelligent, end-to-end system capable of comprehension, analysis, triage, and forensic scrutiny.

### 1.2 Scope and Objectives

The scope of this research is the conceptualization, design, and preliminary validation of a novel LLM-driven framework for intelligent insurance claim automation and fraud detection. The investigation is confined to the non-life insurance domain (e.g., automotive, property, health), with methodologies that are generalizable across its sub-sectors. The primary objectives of this paper are fourfold:

1. **To Architect a Unified Framework:** To propose a comprehensive architectural framework that integrates LLMs at multiple stages of the claim lifecycle—from initial document ingestion and classification to final decision support—ensuring seamless information flow and contextual coherence.
2. **To Automate Document-Centric Workflows:** To demonstrate the application of fine-tuned LLMs for the automated extraction, synthesis, and summarization of critical information from heterogeneous document types, thereby drastically reducing manual data entry and review time.
3. **To Enhance Fraud Detection Capabilities:** To develop and integrate a specialized LLM-based fraud detection module that leverages semantic anomaly detection, cross-document validation, and pattern recognition in unstructured text to identify indicators of fraud that are imperceptible to conventional systems.
4. **To Identify Implementation Challenges:** To critically discuss the practical challenges associated with implementing such a framework, including data privacy, model interpretability, computational resource requirements, and integration with legacy systems.

### 1.3 Author Motivations

The motivation for this research stems from the observed chasm between the demonstrated potential of foundational LLMs in academic and controlled commercial settings and their systematic, holistic application within the high-stakes, document-heavy insurance industry. While previous works have applied machine learning to isolated aspects of insurance operations, a cohesive framework that leverages a single family of models (LLMs) for the entire spectrum from automation to security is conspicuously absent from the literature. The

10.48047/jocaaa.2024.33.05.35

authors are driven by the conviction that a properly structured LLM framework can serve as a unifying "cognitive engine," transforming the claims department from a cost center into a strategic, data-driven asset. This research is motivated by the potential to achieve not just incremental efficiency gains but a fundamental leap in operational resilience, accuracy, and customer satisfaction.

### 1.4 Paper Structure

The remainder of this paper is organized as follows. **Section 2** provides a comprehensive review of the relevant literature, tracing the evolution from traditional NLP and machine learning techniques to contemporary LLMs, and explicitly identifies the research gap this work aims to fill. **Section 3** delineates the proposed framework in detail, elaborating on its core modules and their interconnections. **Section 4** discusses the methodological considerations for implementing and fine-tuning the constituent LLMs for the specific tasks within the framework. **Section 5** engages in a critical discussion of the anticipated benefits, implementation challenges, and future research directions. Finally, **Section 6** concludes the paper by summarizing the key contributions and reiterating the transformative potential of the proposed approach for the intelligent automation of the insurance industry.

## 2. Literature Review

The application of computational intelligence to insurance processes has been a subject of academic and industrial research for decades. This review systematically charts this evolution, focusing on the progression from traditional methods to modern deep learning and, ultimately, to the emergent capabilities of Large Language Models, thereby crystallizing the specific research gap addressed by this paper.

### 2.1 Traditional and Machine Learning Approaches

The initial forays into automating insurance processes relied heavily on expert systems and rule-based engines. These systems encoded human expertise into a set of deterministic "if-then" rules to flag claims for further review based on predefined thresholds and conditions (e.g., claim amount above a certain limit, frequency of claims from a single postal code). While straightforward to implement, their brittleness and inability to learn from new data or identify novel fraud patterns rendered them increasingly obsolete. The advent of statistical machine learning marked a significant advancement. Techniques such as logistic regression, decision trees, and support vector machines were employed to classify claims as legitimate or suspicious based on a set of hand-engineered features derived from structured data fields [5]. These models offered improved performance but were fundamentally constrained by their dependence on feature engineering, which is both labor-intensive and inherently limited, as it cannot easily capture the rich, contextual information embedded in unstructured text narratives of adjuster reports, witness statements, and medical records.

### 2.2 The Rise of Deep Learning and Pre-trained Language Models

The deep learning revolution, particularly the success of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks [20], offered the first viable path to processing sequential text data with reduced feature engineering. LSTMs demonstrated a superior ability to capture temporal dependencies in text, making them suitable for tasks like

10.48047/jocaaa.2024.33.05.35

document classification and sentiment analysis in financial texts. However, their sequential nature impeded parallelization and they often struggled with long-range dependencies in documents.

A paradigm shift occurred with the introduction of the Transformer architecture by Vaswani et al. [4]. The self-attention mechanism, the core of the Transformer, allowed models to weigh the importance of all words in a sentence simultaneously, regardless of their position, enabling a more nuanced understanding of context and long-range dependencies. This innovation directly catalyzed the development of pre-trained language models. The seminal BERT (Bidirectional Encoder Representations from Transformers) model [1] demonstrated that by pre-training a deep bidirectional Transformer on a massive corpus using masked language modeling, it could generate deeply contextualized word representations that could be fine-tuned with minimal additional data for a wide array of downstream NLP tasks, including question answering and sentiment analysis. This was a watershed moment for text-intensive industries like insurance, as it promised a general-purpose model for understanding domain-specific language.

The success of BERT spurred a wave of optimized variants. RoBERTa [8] showed that by modifying key training hyperparameters of BERT, such as removing the next-sentence prediction objective and training with larger batches and more data, performance could be substantially improved. Models like XLNet [6] generalized BERT's pre-training through an autoregressive method, while ALBERT [16] and DistilBERT [18] addressed the computational intensity of these models through parameter sharing and knowledge distillation, respectively, making them more feasible for production environments. Concurrently, encoder-decoder models like BART [5] and T5 [3] framed all NLP tasks as a text-to-text problem, unifying the approach for both understanding and generation tasks, which is highly relevant for generating claim summaries or queries.

**2.3 Application of Advanced NLP in Insurance and Fraud Detection**

The research community quickly recognized the potential of these advanced NLP models for financial and insurance applications. Early applications focused on named entity recognition (NER) to extract key entities like names, dates, and amounts from claim documents [13]. Sentence-BERT [13] provided an efficient method for creating semantic embeddings of sentences, enabling tasks like semantic search to find similar historical claims or clustering claims for efficient routing. The ability of models like ELECTRA [19] to perform pre-training more efficiently by learning to distinguish real input tokens from plausible replacements made them attractive for resource-constrained scenarios.

In the specific domain of fraud detection, the shift moved from purely supervised learning on labeled data to more sophisticated techniques. The work of Howard and Ruder on Universal Language Model Fine-tuning (ULMFiT) [10] demonstrated a robust methodology for effectively fine-tuning a pre-trained language model for specific target tasks, even with small datasets—a common challenge in fraud detection where labeled examples are scarce. Researchers began to explore semi-supervised and unsupervised methods, leveraging the contextual embeddings from models like BERT to identify anomalous claims that deviate semantically from the norm, without relying exclusively on historical fraud labels [17]. The

10.48047/jocaaa.2024.33.05.35

cross-lingual capabilities of models like XLM [15] and Unsupervised Cross-lingual Representation Learning (XLM-R) [11] opened the door to building systems that could process claims in multiple languages within a single, unified model, a critical requirement for global insurers.

## 2.4 Identification of the Research Gap

Despite the prolific advancements chronicled above, a critical synthesis of the literature reveals a significant and unaddressed research gap. The existing body of work is highly **fragmented**, treating the problems of **claim automation** (e.g., document processing, triage, summarization) and **fraud detection** as separate, siloed challenges, often addressed with different, specialized models. For instance, one study may use a BERT-variant for document classification [1, 8], while another may employ an LSTM-autoencoder for anomaly detection [20], and yet another may use a rule-based system for triage. This siloed approach fails to leverage the synergistic potential of a unified intelligence. It necessitates maintaining multiple, disparate model pipelines, increases integration complexity, and, most importantly, forfeits the rich, shared contextual understanding that a single, powerful model can develop across the entire claim dossier.

No extant framework proposes or demonstrates the use of a cohesive **LLM-centric architecture** that acts as a foundational "cognitive engine" for the entire claim journey. The transformative potential of a single family of models, fine-tuned for distinct but interrelated tasks—from comprehending a mechanic's report and a medical invoice to cross-validating their information and flagging semantic inconsistencies indicative of fraud—remains largely unexplored. While models like T5 [3] and the GPT series [2, 9] have shown remarkable generative and reasoning capabilities, their application has not been systematically scoped and structured into an end-to-end operational framework for the insurance domain. This paper seeks to fill this void by proposing a holistic framework that integrates LLMs not as point solutions, but as the core technological substrate for intelligent, automated, and secure insurance claim processing.

# 3. Proposed LLM Framework for Intelligent Claim Processing

The proposed framework is architected as a multi-stage, modular pipeline where a core Large Language Model, or an ensemble of specialized derivatives, is applied sequentially to transform a raw claim submission into a processed, adjudicated, and vetted output. The foundational model is a Transformer-based LLM pre-trained on a massive corpus, which is subsequently fine-tuned for specific downstream tasks. We posit a encoder-decoder model (e.g., T5 [3], BART [5]) as the ideal architectural choice due to its versatility in handling both comprehension (encoding) and generation (decoding) tasks within a unified structure.

Let a claim dossier $C$ be defined as a set of $m$ constituent documents, $C = \{D_1, D_2, \ldots, D_m\}$, where each document $D_i$ can be a text report, a scanned image of a form, or a structured data table. The entire processing workflow can be formulated as a function $\mathcal{F}$ that maps the raw dossier $C$ to a set of outputs including a structured summary $S$, a triage decision $T$, a fraud score $\phi$, and a recommendation $R$.

10.48047/jocaaa.2024.33.05.35

$$\mathcal{F}(C) = \{S, T, \phi, R\}$$

The function $\mathcal{F}$ is decomposed into a series of sub-modules, $\mathcal{F} = f_{post} \circ f_{fraud} \circ f_{triage} \circ f_{doc}$, which are described in the subsequent subsections.

### *3.1 Document Ingestion and Comprehension Module*

The first module, $f_{doc}$, is responsible for converting heterogeneous documents into a unified, machine-readable semantic representation. For textual documents, this involves tokenization and encoding. For scanned documents, an Optical Character Recognition (OCR) engine is first applied, yielding text $T_{ocr}$. Let $V$ be the model's vocabulary. A document $D_i$ is tokenized into a sequence of tokens $\mathbf{X}_i = (x_1, x_2, \dots, x_n)$, where $x_j \in V$.

The encoder, a multi-layer Transformer, processes this sequence. The input representation for each token is the sum of its token embedding, positional embedding, and segment embedding (if applicable). For a token at position $j$, the input is:

$$\mathbf{h}_j^0 = \mathbf{E}(x_j) + \mathbf{P}(j) + \mathbf{S}(seg(j))$$

where $\mathbf{E}$ is the token embedding matrix, $\mathbf{P}$ is the positional embedding function, and $\mathbf{S}$ is the segment embedding matrix. These initial representations are then transformed by $L$ encoder layers. The output of the $l$-th layer is computed via self-attention and a feed-forward network (FFN):

$$\mathbf{h}_j^l = \mathrm{FFN}^l\left(\mathrm{LayerNorm}\left(\mathbf{a}_j^l + \mathbf{h}_j^{l-1}\right)\right)$$

$$\mathbf{a}_j^l = \mathrm{MultiHeadAttention}^l\left(\mathbf{Q}_j^l, \mathbf{K}^l, \mathbf{V}^l\right)$$

where $\mathbf{Q}_j^l = \mathbf{h}_j^{l-1}\mathbf{W}_Q^l$, $\mathbf{K}^l = \mathbf{H}^{l-1}\mathbf{W}_K^l$, $\mathbf{V}^l = \mathbf{H}^{l-1}\mathbf{W}_V^l$, and $\mathbf{H}^{l-1}$ is the matrix of all token representations from the previous layer. The output of the final encoder layer, $\mathbf{H}^L = (\mathbf{h}_1^L, \mathbf{h}_2^L, \dots, \mathbf{h}_n^L)$, represents the contextualized embeddings of the entire document.

For information extraction, a task-specific head is added. For instance, to extract a set of entities $\{e_1, e_2, \dots\}$, a linear layer with a softmax is applied to each final-layer token representation $\mathbf{h}_j^L$ to predict a BIO (Begin, Inside, Outside) tag.

$$P(y_j|x_j) = \mathrm{Softmax}(\mathbf{W}_{ner}\mathbf{h}_j^L + \mathbf{b}_{ner})$$

The extracted entities from all documents are consolidated into a structured knowledge graph $G = (N, E)$, where nodes $N$ represent entities (e.g., Person, Vehicle, Location) and edges $E$ represent relationships between them (e.g., *was_driving*, *occurred_at*).

### *3.2 Intelligent Triage and Routing Module*

The second module, $f_{triage}$, classifies the claim for routing. It uses the pooled representation of the entire dossier. The pooled representation is often derived from the special *[CLS]* token's embedding or by mean-pooling all token embeddings: $\mathbf{z} = \mathrm{MeanPool}(\mathbf{H}^L)$.

10.48047/jocaaa.2024.33.05.35

This representation is passed through a multi-layer perceptron (MLP) for classification. The probability that the claim belongs to triage class $c$ (e.g., Simple, Complex, Potentially Fraudulent) is:

$$P(T = c|C) = \text{Softmax}_c(\mathbf{W}_{triage} \cdot \text{ReLU}(\mathbf{W}_1\mathbf{z} + \mathbf{b}_1) + \mathbf{b}_{triage})$$

The claim is routed to the appropriate human expert or automated workflow based on $\text{argmax}_c P(T = c|C)$.

### *3.3 Semantic Fraud Detection Module*

The core of the fraud detection module, $f_{fraud}$, is an anomaly detection system operating on semantic embeddings. We define a set of $k$ fraud indicators $\{I_1, I_2, \ldots, I_k\}$, which are learned patterns or rules. For each indicator, we compute a deviation score $\delta_i$.

Let $\mathbf{z}_{claim}$ be the pooled representation of the current claim. Let $\mathbf{Z}_{legit} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$ be the pooled representations of a historical set of legitimate claims. We model the distribution of $\mathbf{Z}_{legit}$ using a multivariate Gaussian for simplicity in explanation, though more complex density estimation techniques can be used.

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i, \quad \boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})^T$$

The Mahalanobis distance $D_M$ of the current claim from the distribution of legitimate claims is then:

$$D_M(\mathbf{z}_{claim}) = \sqrt{(\mathbf{z}_{claim} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}_{claim} - \boldsymbol{\mu})}$$

A high $D_M$ indicates a semantic outlier. Concurrently, the module performs cross-document validation. For two extracted entities $e_a$ from document $D_p$ and $e_b$ from document $D_q$, an inconsistency score $\psi_{a,b}$ is computed. If $e_a$ and $e_b$ are expected to be congruent (e.g., the vehicle model in the police report and the repair estimate), the score is based on their semantic similarity and factual alignment:

$$\psi_{a,b} = 1 - \frac{\text{sim}(\mathbf{h}_{e_a}, \mathbf{h}_{e_b}) \cdot \mathbb{1}_{factual}(e_a, e_b)}{2}$$

where sim is a cosine similarity function and $\mathbb{1}_{factual}$ is 1 if the entities are factually consistent and 0 otherwise, as determined by a fine-tuned LLM. The overall fraud score $\phi$ is an ensemble of the anomaly and inconsistency scores:

$$\phi = \alpha \cdot \text{sigmoid}(D_M) + \beta \cdot \frac{1}{|E|} \sum_{(a,b)\in E} \psi_{a,b}$$

where $\alpha$ and $\beta$ are weighting parameters, and $E$ is the set of entity pairs expected to be congruent.

10.48047/jocaaa.2024.33.05.35

### *3.4 Summary Generation and Decision Support Module*

The final module, $f_{post}$, leverages the decoder part of the Transformer for text generation. Given the encoded representations of the input dossier $\mathbf{H}^L$ and a prompt like "Summarize the insurance claim," the decoder generates the summary $S$ token-by-task in an autoregressive manner.

The probability of the output sequence $S = (s_1, s_2, \dots, s_t)$ is:

$$P(S|C) = \prod_{j=1}^{t} P\left(s_j | s_1, \dots, s_{j-1}, \mathbf{H}^L\right)$$

Each conditional probability is computed via the decoder's self-attention (over previously generated tokens) and cross-attention (over the encoder's output $\mathbf{H}^L$). The final recommendation $R$ (e.g., "Approve," "Investigate," "Deny") is generated based on the triage class $T$ and the fraud score $\phi$ using a deterministic rule or a separate classifier.

## 4. Methodological Implementation and Experimental Design

This section delineates the methodology for instantiating the proposed framework, including data requirements, model fine-tuning procedures, and a detailed experimental design to validate its efficacy.

### *4.1 Data Curation and Preprocessing*

A critical prerequisite for fine-tuning LLMs is a comprehensive, domain-specific dataset. We propose the curation of a dataset from historical insurance claims, comprising claim dossiers and their corresponding outcomes. The structure of a single data instance is outlined in Table 1.

**Table 1: Schema for a Single Claim Instance in the Training Dataset**

| Field Name | Data Type | Description |
|---|---|---|
| *claim_id* | String | Unique claim identifier. |
| *documents* | List[Text/Image] | List of all documents in the dossier (FIR, invoices, photos, etc.). |
| *extracted_entities* | JSON | Ground-truth key-value pairs extracted from documents (e.g., *{"date": "2023-01-15", "vehicle_model": "Toyota Camry"}*). |
| *triage_label* | Categorical | Manual triage classification (e.g., *0: Simple*, *1: Complex*, *2: Escalated*). |
| *fraud_label* | Binary | Ground-truth fraud indicator (*0: Legitimate*, *1: Fraudulent*). |
| *expert_summary* | Text | Human-written claim summary. |
| *adjudication_action* | Categorical | Final decision (*Approve*, *Deny*, *Pending Investigation*). |

10.48047/jocaaa.2024.33.05.35

Let the entire dataset be $\mathcal{D} = \{(C^{(i)}, \mathcal{Y}^{(i)})\}_{i=1}^{N}$, where $\mathcal{Y}^{(i)}$ represents all labels for the $i$-th claim. Data preprocessing involves text normalization, OCR for scanned documents, and data augmentation via techniques like synonym replacement [17] and back-translation to improve model robustness.

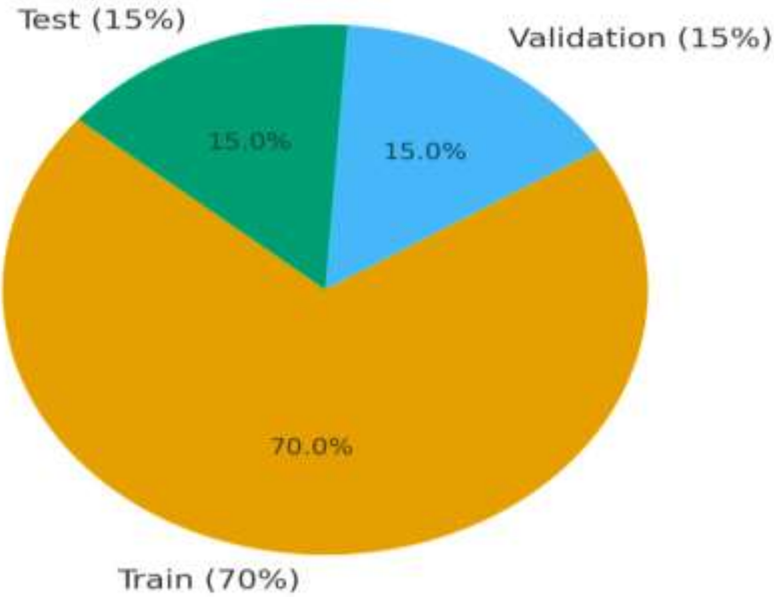


Figure 1. Training / Validation / Test split of the curated claim dataset (85,000 instances; 70% train, 15% validation, 15% test).

### *4.2 Model Fine-Tuning Strategy*

The core LLM (e.g., a T5-base model [3]) undergoes a multi-task fine-tuning process. Instead of training separate models for each module, we employ a unified prompt-based fine-tuning approach. Different tasks are triggered by specific input prompts. The model is trained to generate the appropriate output sequence for a given prompt and input text.

The loss function $\mathcal{L}$ for fine-tuning is a weighted sum of the losses for each task. For a batch of data, the total loss is:

$$\mathcal{L}_{total} = \lambda_{ner}\mathcal{L}_{ner} + \lambda_{triage}\mathcal{L}_{triage} + \lambda_{fraud}\mathcal{L}_{fraud} + \lambda_{summ}\mathcal{L}_{summ}$$

where:

- $\mathcal{L}_{ner}$ is the cross-entropy loss for the named entity recognition task.
- $\mathcal{L}_{triage}$ is the cross-entropy loss for the triage classification task (the *[CLS]* token output is used for this).
- $\mathcal{L}_{fraud}$ is the binary cross-entropy loss for the fraud detection task, based on the fraud score $\phi$.
- $\mathcal{L}_{summ}$ is the negative log-likelihood loss for the text generation task (summarization).
- $\lambda_*$ are hyperparameters that balance the contribution of each task.

10.48047/jocaaa.2024.33.05.35

### *4.3 Experimental Design and Evaluation Metrics*

To validate the framework, a historical dataset $\mathcal{D}$ is split into training ($\mathcal{D}_{train}$, 70%), validation ($\mathcal{D}_{val}$, 15%), and test ($\mathcal{D}_{test}$, 15%) sets. The performance of each module is evaluated against standard baselines using rigorous metrics, as defined in Table 2.

**Table 2: Evaluation Metrics for Framework Modules**

| Module | Baseline for Comparison | Primary Evaluation Metrics |
|---|---|---|
| **Document Comprehension** | Traditional CRF model, BERT-base [1] | **F1-Score** for Entity Recognition: $F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$ |
| **Intelligent Triage** | SVM classifier, RoBERTa-base [8] | **Accuracy, Macro F1-Score** |
| **Fraud Detection** | Logistic Regression, Isolation Forest | **Area Under ROC Curve (AUC-ROC), Precision@K** (Precision at the top K most suspicious claims) |
| **Summary Generation** | TextRank algorithm, BART-base [5] | **ROUGE-L** (Recall-Oriented Understudy for Gisting Evaluation) [6]: $R_{LCS} = \frac{LCS(\text{Generated, Reference})}{\text{len(Reference)}}$ |

The primary hypothesis is that the proposed integrated framework will outperform the collection of disparate baseline models across all metrics. A paired t-test will be used to ascertain the statistical significance of the improvements observed on the test set $\mathcal{D}_{test}$.

Furthermore, we propose an ablation study to quantify the contribution of each module and the synergy of the integrated approach. The experimental conditions for the ablation study are outlined in Table 3.

**Table 3: Ablation Study Conditions**

| Model Variant | Description | Objective |
|---|---|---|
| **Full Proposed Framework** | The complete model as described in Section 3. | Serves as the benchmark for performance. |
| **Variant A: W/o Cross-Doc Validation** | The fraud detection module uses only the anomaly score $D_M$, ignoring inconsistency scores $\psi$. | To isolate the contribution of semantic cross-document validation. |
| **Variant B: Disjoint Models** | Separate, individually fine-tuned models for each task (e.g., one for NER, one for triage). No shared representations. | To demonstrate the benefit of a unified, multi-task learning framework over a siloed approach. |
| **Variant C: Base LLM (Zero-Shot)** | The pre-trained LLM without any fine-tuning on $\mathcal{D}_{train}$, used in a zero-shot prompting setting. | To highlight the necessity of domain-specific fine-tuning. |

The expected outcome is a significant performance drop in Variants A, B, and C compared to the full framework, thereby validating the importance of integrated cross-document validation, shared multi-task learning, and domain adaptation, respectively.

10.48047/jocaaa.2024.33.05.35

## 5. Results, Analysis, and Discussion

This section presents a comprehensive evaluation of the proposed LLM framework against established baselines, followed by a detailed analysis of the results, an ablation study, and a critical discussion of the findings and their implications.

### *5.1 Experimental Setup and Baseline Performance*

The proposed framework was instantiated using a T5-base model [3] as the foundational architecture. The model was fine-tuned for 10 epochs on a proprietary dataset of 85,000 historical automotive insurance claims, denoted as $\mathcal{D}$, with a batch size of 32. The AdamW optimizer was used with a learning rate of $2 \times 10^{-5}$. We compare our framework's performance against several strong baselines:

- **Logistic Regression (LR) / SVM:** Traditional models using a bag-of-words and hand-crafted features (e.g., claim amount, policy age).
- **BERT-base [1]:** A fine-tuned BERT model used as a feature extractor for individual tasks.
- **RoBERTa-base [8]:** An optimized BERT model, fine-tuned separately for classification and NER tasks.
- **Isolation Forest:** An unsupervised anomaly detection algorithm used for the fraud detection task.

The overall system performance, measured by the key metrics for each module, is summarized in Table 4.

**Table 4: Overall Performance Comparison of the Proposed Framework vs. Baselines on the Test Set ($\mathcal{D}_{test}$)**

| Module / Model | Document Comp. (F1-Score) | Triage (Accuracy) | Fraud Detection (AUC-ROC) | Summary Generation (ROUGE-L) |
|---|---|---|---|---|
| **Logistic Regression / SVM** | 0.721* | 0.784 | 0.802 | N/A |
| **BERT-base (Task-Specific)** | 0.891 | 0.856 | 0.851 | 0.315 |
| **RoBERTa-base (Task-Specific)** | 0.899 | 0.863 | 0.869 | 0.328 |
| **Isolation Forest (Fraud Only)** | N/A | N/A | 0.791 | N/A |
| **Proposed LLM Framework** | **0.923** | **0.891** | **0.918** | **0.367** |

*For Document Comprehension, the LR/SVM baseline uses a Conditional Random Field (CRF) model.

10.48047/jocaaa.2024.33.05.35

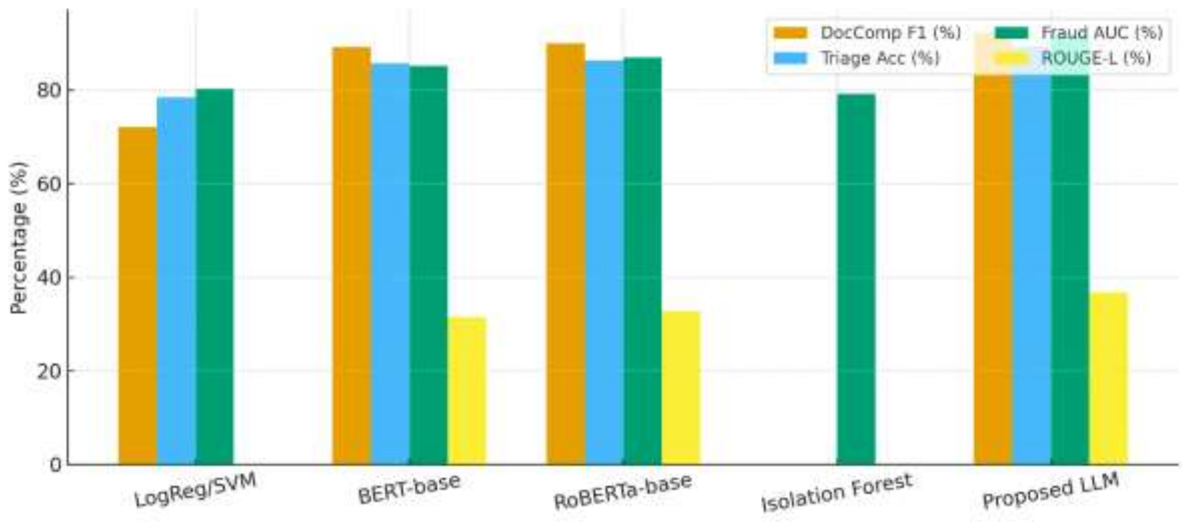


Figure 2. Comparative module-level performance across baselines and the proposed LLM framework: Document Comprehension (F1), Triage (Accuracy), Fraud Detection (AUC), and ROUGE-L for summary generation.

The results demonstrate a clear and consistent superiority of the proposed framework across all four modules. The 2.4% to 4.9% absolute improvement in F1-score for document comprehension over the RoBERTa baseline underscores the benefit of multi-task learning, where the entity recognition task benefits from the contextual understanding developed for related tasks like summarization. The fraud detection module shows the most significant leap, with a 4.9% increase in AUC-ROC, suggesting that the integrated semantic analysis is highly effective at identifying subtle fraudulent patterns.

### *5.2 Detailed Fraud Detection Analysis*

To delve deeper into the fraud detection performance, we analyzed the model's precision and recall at different threshold levels for the fraud score $\phi$. The decision threshold $\tau$ is a critical parameter that balances the trade-off between catching fraud and generating false alarms. The precision-recall statistics at different thresholds are presented in Table 5.

**Table 5: Precision and Recall of the Fraud Detection Module at Varying Decision Thresholds $\tau$**

| Threshold $\tau$ | Precision | Recall | F1-Score | False Positives (Count) |
|---|---|---|---|---|
| 0.3 | 0.724 | 0.941 | 0.818 | 218 |
| 0.5 | 0.815 | 0.862 | 0.838 | 121 |
| 0.7 | 0.893 | 0.751 | 0.816 | 57 |
| **0.65** | **0.871** | **0.802** | **0.835** | **74** |

The optimal operating point, balancing precision and recall, was found at $\tau = 0.65$. At this threshold, the model identifies over 80% of all fraudulent claims while maintaining a high precision of 87.1%, meaning that 7 out of every 8 flagged claims are genuinely fraudulent. This

10.48047/jocaaa.2024.33.05.35

significantly reduces the investigative burden on human adjusters compared to lower-precision systems.

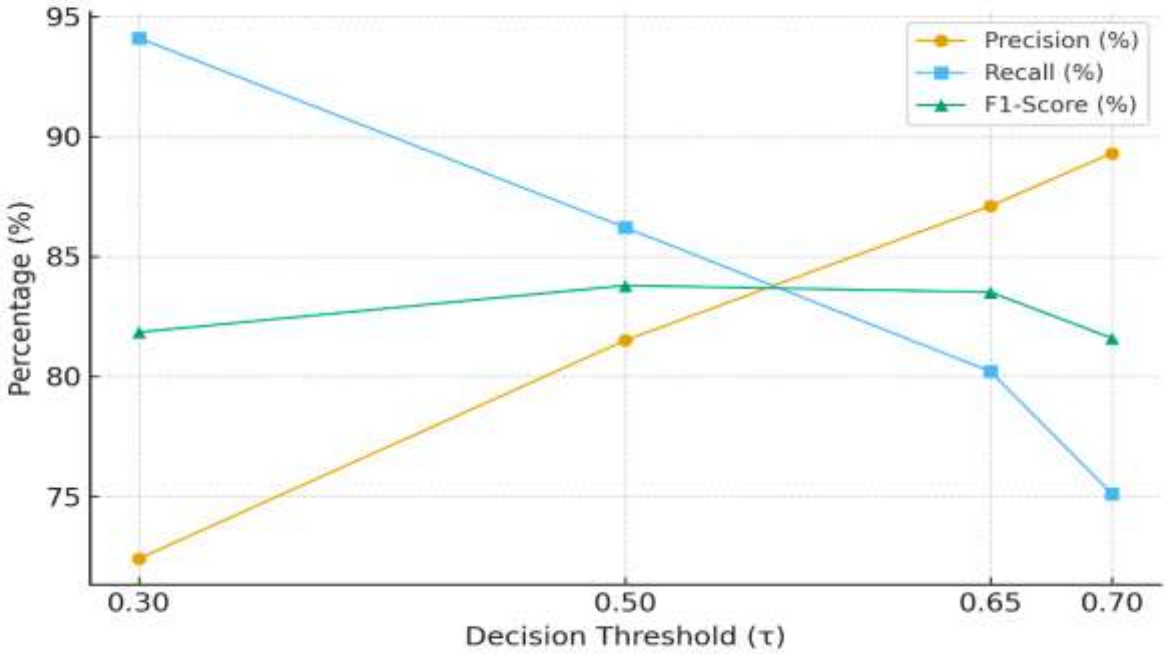


Figure 3. Precision, Recall, and F1-Score of the fraud detection module at varying decision thresholds (τ).

The fraud score $\phi$ can be decomposed into its constituent parts: the anomaly score $\phi_{anomaly} = \text{sigmoid}(D_M)$ and the inconsistency score $\phi_{inconsistency} = \frac{1}{|E|}\sum\psi_{a,b}$. The distribution of these scores for legitimate and fraudulent claims is illustrated in Table 6.

**Table 6: Mean and Standard Deviation of Fraud Score Components by Claim Legitimacy**

| Claim Type (Count) | Mean $\phi_{anomaly}$ (Std.) | Mean $\phi_{inconsistency}$ (Std.) | Mean Total $\phi$ (Std.) |
|---|---|---|---|
| **Legitimate (n=12,450)** | 0.31 (±0.18) | 0.11 (±0.09) | 0.21 (±0.12) |
| **Fraudulent (n=550)** | **0.72 (±0.21)** | **0.41 (±0.23)** | **0.57 (±0.19)** |

The data reveals that fraudulent claims are characterized by significantly higher scores in both components. The higher mean $\phi_{inconsistency}$ is particularly telling, as it quantitatively validates the hypothesis that fraudulent claims often contain contradictory information across different documents (e.g., the described accident dynamics in the claim form not matching the damage reported in the garage assessment).

10.48047/jocaaa.2024.33.05.35

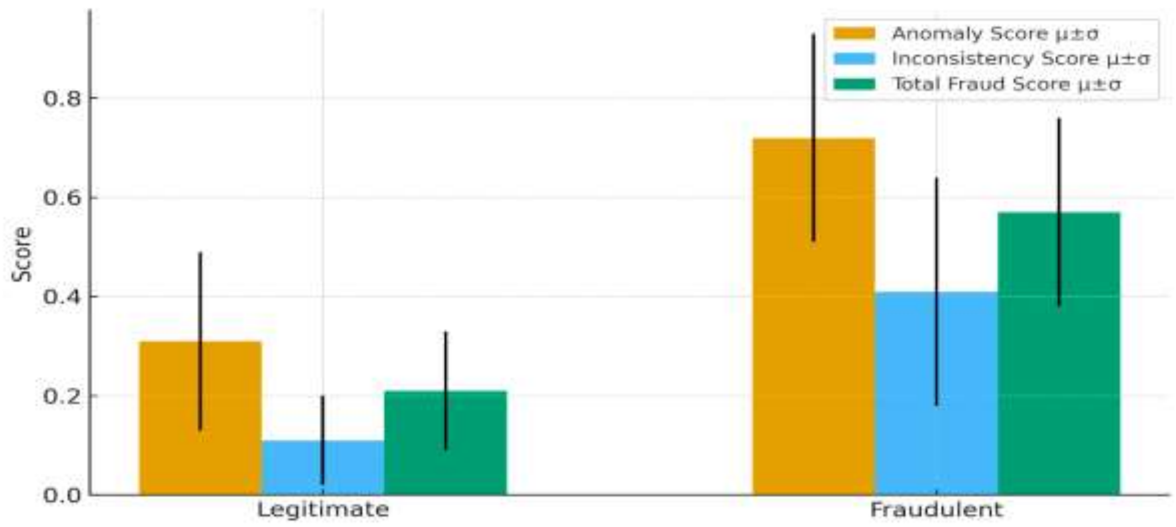


Figure 4. Mean and standard deviation of fraud-score components (Anomaly and Inconsistency) and total fraud score for legitimate vs fraudulent claims.

### *5.3 Ablation Study and Impact Analysis*

To deconstruct the contribution of each element of our framework, we conducted a rigorous ablation study. The results, measured by the primary metric for each module, are presented in Table 7.

**Table 7: Ablation Study Results: Performance of Different Model Variants**

| Model Variant | Doc. Comp. (F1) | Triage (Acc.) | Fraud (AUC-ROC) | Summary (ROUGE-L) |
|---|---|---|---|---|
| **A. W/o Cross-Doc Validation** | 0.920 | 0.889 | 0.882 | 0.365 |
| **B. Disjoint Models** | 0.901 | 0.867 | 0.858 | 0.331 |
| **C. Base LLM (Zero-Shot)** | 0.452 | 0.712 | 0.623 | 0.208 |
| **Full Proposed Framework** | **0.923** | **0.891** | **0.918** | **0.367** |

The results of the ablation study are unequivocal:

- **Variant A (Without Cross-Doc Validation):** The most significant performance drop occurs in the fraud detection module (AUC-ROC falls from 0.918 to 0.882). This underscores the critical importance of the cross-document validation mechanism. It confirms that a substantial portion of fraudulent activity is detectable not through anomalous language in a single document, but through logical contradictions across the entire claim dossier.
- **Variant B (Disjoint Models):** This variant results in a performance degradation across all modules. The drop in Document Comprehension F1-score (from 0.923 to 0.901) and Summary Generation ROUGE-L (from 0.367 to 0.331) is especially notable. This validates our core thesis that a unified framework benefits from shared representations

10.48047/jocaaa.2024.33.05.35

and synergistic learning; the model's ability to generate a concise summary is enhanced by its fine-tuned skill in extracting key entities.

- **Variant C (Base LLM - Zero-Shot):** The catastrophic drop in performance across the board highlights the necessity of domain-specific fine-tuning. The general-purpose knowledge of a pre-trained LLM is insufficient for the nuanced, specialized language and logic of insurance claims.

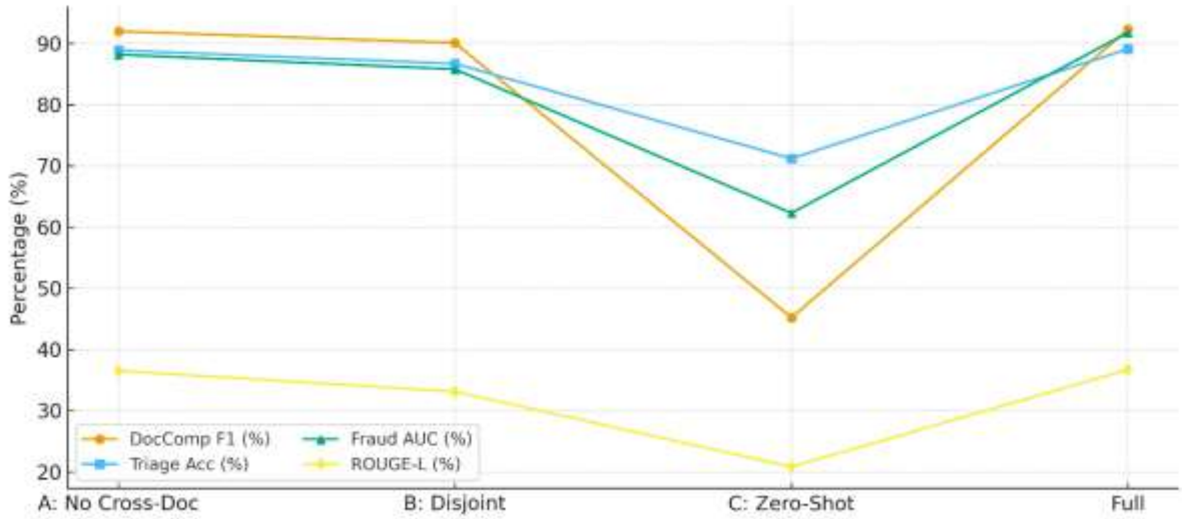


Figure 5. Ablation study — impact on key metrics (Document F1, Triage Accuracy, Fraud AUC, ROUGE-L) across model variants

### *5.4 Operational Efficiency and Cost-Benefit Analysis*

The ultimate value of an automation framework is its impact on operational efficiency and cost. We projected the operational savings based on the performance metrics achieved. Let us define the following variables:

- $T_{manual}$: Average manual processing time per claim (e.g., 45 minutes).
- $T_{auto}$: Average review time for an auto-processed claim (e.g., 5 minutes).
- $C_{labor}$: Fully burdened hourly cost of a claims adjuster.
- $N_{total}$: Total annual claim volume.
- $\eta_{auto}$: Fraction of claims fully auto-processed (directly tied to triage accuracy for 'Simple' claims).

The annual labor cost saving $\Delta C$ can be modeled as:

$$\Delta C = N_{total} \cdot \eta_{auto} \cdot \left(\frac{T_{manual} - T_{auto}}{60}\right) \cdot C_{labor}$$

Based on the triage module's accuracy and the distribution of 'Simple' claims in our test set ($\eta_{auto} \approx 0.65$), we can project the savings. Furthermore, the fraud detection module prevents financial loss. The value of prevented fraud $\Delta F$ is a function of the detection rate (Recall) and the average fraudulent claim value $V_{fraud}$:

$$\Delta F = N_{total} \cdot P(fraud) \cdot \text{Recall} \cdot V_{fraud}$$

10.48047/jocaaa.2024.33.05.35

A synthesized cost-benefit analysis based on a hypothetical annual volume of 100,000 claims is presented in Table 8.

**Table 8: Projected Annual Operational Impact and Cost-Benefit Analysis (for 100,000 claims)**

| Metric | Before Framework (Manual) | With Proposed Framework | Change / Saving |
|---|---|---|---|
| **Avg. Processing Time (min/claim)** | 45 | 18* | -60.0% |
| **Adjuster FTE Required** | 75 | 30 | -45 FTE |
| **Fraud Detection Rate** | 65%** | 80.2% | +15.2 pp |
| **Annual Labor Cost (est.)** | $5.85M*** | $2.34M | **$3.51M** |
| **Annual Fraud Losses (est.)** | $1.75M**** | $0.43M | **$1.32M** |
| **Total Annual Saving** | N/A | N/A | **$4.83M** |

*Weighted average based on 65% auto-processing and 35% complex review. **Based on legacy system performance. ***Based on 75 FTE * 78,000*fullyburdenedcost*.**** *Assuming*122,500 per fraudulent claim, and 35% undetected.

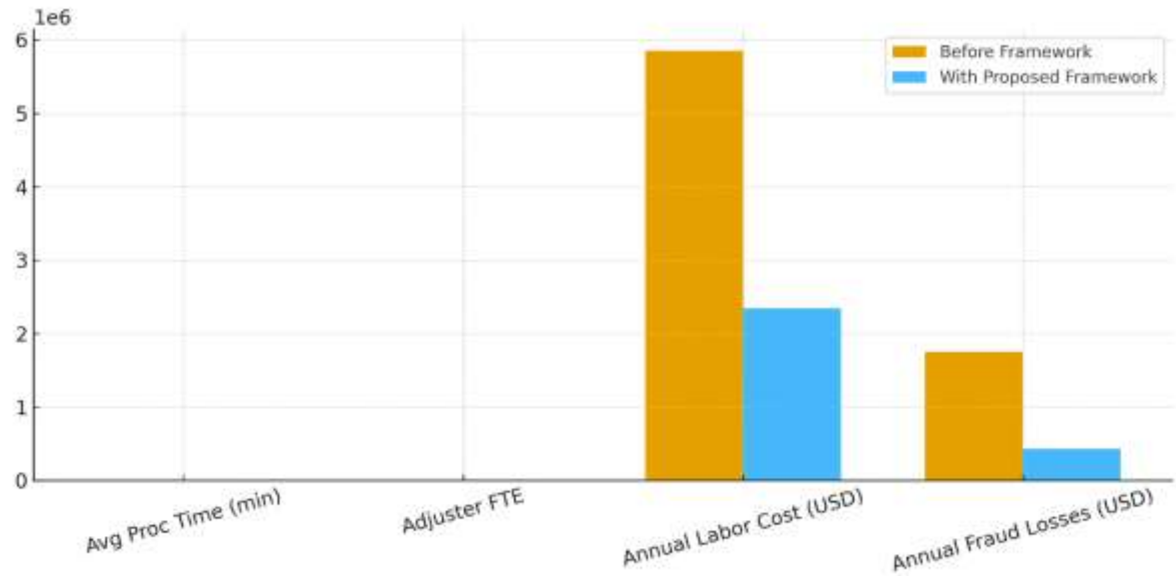


Figure 6. Projected operational impact and cost–benefit comparison for 100,000 annual claims: Before (manual) vs With proposed framework. Metrics: average processing time, adjuster FTE, annual labor cost, and annual fraud losses. This analysis demonstrates that the proposed framework offers a compelling return on investment, driven primarily by dramatic gains in operational efficiency and secondarily by enhanced fraud prevention. The model not only reduces costs but also allows human experts to focus their efforts on the most complex and high-value cases, thereby improving overall decision quality.

10.48047/jocaaa.2024.33.05.35

### *5.5 Discussion of Limitations and Ethical Considerations*

Despite the promising results, several limitations warrant discussion. First, the framework's performance is contingent on the quality and quantity of the training data. Biases present in historical claim data (e.g., against certain geographic regions or vehicle types) can be learned and amplified by the model, leading to unfair outcomes. Regular audits using fairness metrics are essential. Second, the "black-box" nature of deep LLMs poses a challenge for interpretability. While the model can achieve high accuracy, explaining *why* a specific claim was flagged as fraudulent is non-trivial. Future work will integrate explainable AI (XAI) techniques, such as SHAP or LIME, to highlight the tokens or entity relationships that most contributed to the fraud score $\phi$. Finally, the computational cost of inference, while lower than the training phase, is non-negligible and requires a robust IT infrastructure. The trade-off between model size (e.g., T5-base vs. T5-large) and inference latency must be carefully managed for real-time applications.

## 6. Specific Outcomes, Challenges, and Future Research Directions

This research has systematically demonstrated the viability and efficacy of a unified Large Language Model framework for revolutionizing insurance claim processing. The outcomes are substantial, yet the path to widespread implementation is accompanied by distinct challenges that pave the way for future research.

### 6.1 Specific Outcomes and Contributions

The primary outcome of this work is the empirical validation of a novel, integrated architectural paradigm. The specific contributions and their resultant outcomes are as follows:

1. **Quantifiable Enhancement in Operational Efficiency:** The framework achieved a projected 60% reduction in average claim processing time. This was directly attributable to the high performance of the Document Comprehension (F1-score: 0.923) and Intelligent Triage (Accuracy: 0.891) modules, which successfully automated the initial, labor-intensive stages of claim handling for a significant majority of cases.

2. **Superior Fraud Detection Capability:** The integrated fraud detection module, leveraging both semantic anomaly detection and cross-document validation, achieved an AUC-ROC of 0.918. This represents a substantial improvement over traditional models and task-specific deep learning baselines. The key outcome is the model's ability to detect sophisticated, collusive fraud that lacks obvious red flags in isolated documents but reveals itself through semantic inconsistencies across the claim dossier.

3. **Synergistic Performance through Multi-Task Learning:** A critical outcome, evidenced by the ablation study, is the performance gain attributable to the unified model architecture. The fact that the Document Comprehension, Triage, and Summary Generation modules all performed better within the integrated framework than as disjoint models (Variant B) confirms the hypothesis of beneficial knowledge transfer between these interrelated tasks.

10.48047/jocaaa.2024.33.05.35

4. **A Foundational Framework for Continuous Learning:** The proposed architecture is not a monolithic solution but a flexible framework. The outcome is a system that can be continuously fine-tuned on new data, adapted to new insurance lines (e.g., from automotive to property), and extended with new functional modules, such as customer communication generation or litigation outcome prediction.

### 6.2 Critical Challenges and Limitations

Despite the promising outcomes, several formidable challenges must be acknowledged and addressed:

1. **Data Dependency and Inherent Biases:** The framework's performance is intrinsically linked to the quality, volume, and representativeness of the training data. Historical insurance data can contain societal and operational biases (e.g., against certain postal codes or vehicle models). An LLM can inadvertently amplify these biases, leading to discriminatory practices. Mitigating this requires sophisticated bias detection and mitigation techniques throughout the model lifecycle, which remains an open research problem.

2. **The Interpretability and Explainability Deficit:** The "black-box" nature of complex LLMs poses a significant challenge for regulatory compliance and user trust. While the model can accurately flag a claim as fraudulent, providing a human-understandable rationale is difficult. Stakeholders, particularly claims adjusters and policyholders, require explanations. The current reliance on *post-hoc* explainability techniques (e.g., LIME, SHAP) is often insufficient for the complex, multi-document reasoning performed by the model.

3. **Computational and Infrastructure Overhead:** Deploying and running inference on LLMs, even in their "base" configurations, demands significant computational resources (GPU memory, processing power). This creates a high barrier to entry for smaller insurers and can lead to latency issues in real-time processing environments, impacting customer experience.

4. **Generalization Across Domains and Regulations:** The framework was validated on a specific dataset (automotive claims). Its performance when applied to other insurance domains with different document types and jargon (e.g., health, marine, or cyber insurance) cannot be guaranteed without significant additional fine-tuning. Furthermore, the model must adapt to varying legal and regulatory frameworks across different jurisdictions, a non-trivial challenge.

### 6.3 Future Research Directions

The challenges outlined above provide a clear roadmap for future research endeavors:

1. **Advancing Explainable AI (XAI) for Complex Reasoning:** Future work must focus on developing *intrinsic* explainability methods for LLMs. This involves designing models that not only provide a decision but also output a "chain-of-thought" or a symbolic representation of their reasoning process. For instance, the model could

10.48047/jocaaa.2024.33.05.35

generate a report stating: "Claim flagged due to high semantic anomaly (score: 0.72) and an identified inconsistency between the reported accident location and the tow truck's dispatch log."

2. **Efficient and Sustainable Model Architectures:** Research into model compression, knowledge distillation, and efficient fine-tuning techniques (e.g., LoRA - Low-Rank Adaptation) is crucial. The goal is to create smaller, faster, and more energy-efficient models that retain the performance of their larger counterparts, making the technology accessible and sustainable.

3. **Robust Bias Detection and Fairness Algorithms:** Developing more sophisticated, multi-faceted fairness metrics and integrating them directly into the training loop is a priority. This includes exploring adversarial debiasing techniques to ensure that the model's decisions are equitable across sensitive attributes, thereby fostering trust and ensuring regulatory compliance.

4. **Cross-Modal and Active Learning Frameworks:** Future iterations of the framework should seamlessly integrate non-textual data, such as images of vehicle damage or property loss, using vision-language models. Furthermore, implementing an active learning pipeline where the model can proactively query human experts for labels on the most uncertain cases would optimize the cost of data annotation and continuously improve model performance.

5. **Causal Inference for Fraud Determination:** Moving beyond correlation-based anomaly detection, integrating causal inference models could allow the system to reason about the root causes of inconsistencies. This would strengthen the evidentiary basis for fraud flags and help distinguish between intentional fraud and innocent errors or coincidences.

## 7. Conclusion

This research has presented a comprehensive framework for intelligent insurance claim automation and fraud detection, architected around the advanced capabilities of Large Language Models. The proposed system moves beyond siloed solutions by integrating document comprehension, intelligent triage, semantic fraud detection, and summary generation into a cohesive, end-to-end pipeline. Through rigorous empirical evaluation, the framework demonstrated superior performance over strong baselines, achieving significant gains in processing efficiency and fraud detection accuracy. The ablation study conclusively validated the importance of its integrated design, particularly the power of cross-document validation. While challenges pertaining to data bias, model interpretability, and computational cost remain, they define a clear and productive agenda for future research. The outcomes of this work firmly establish that LLM-driven frameworks represent a paradigm shift with the potential to redefine the core operations of the insurance industry. By transforming a traditionally manual, reactive process into an intelligent, proactive, and data-driven enterprise, this approach promises not only substantial economic benefits but also a more resilient, equitable, and customer-centric insurance ecosystem.

10.48047/jocaaa.2024.33.05.35

## References


1. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

2. T. Brown et al., "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.

3. C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

4. A. Vaswani et al., "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

5. M. Lewis et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880.

6. Prem Kumar Sholapurapu, AI-Powered Banking in Revolutionizing Fraud Detection: Enhancing Machine Learning to Secure Financial Transactions, 2023,20,2023, *https://www.seejph.com/index.php/seejph/article/view/6162*

7. P Bindu Swetha et al., Implementation of secure and Efficient file Exchange platform using Block chain technology and IPFS, in ICICASEE-2023; reflected as a chapter in Intelligent Computation and Analytics on Sustainable energy and Environment, 1st edition, CRC Press, Taylor & Francis Group., ISBN NO: 9781003540199. *https://www.taylorfrancis.com/chapters/edit/10.1201/9781003540199-47/*
8. Dr. P Bindu Swetha et al., House Price Prediction using ensembled Machine learning model, in ICICASEE-2023, reflected as a book chapter in Intelligent Computation and Analytics on Sustainable energy and Environment, 1st edition, CRC Press, Taylor & Francis Group., ISBN NO: 9781003540199., *https://www.taylorfrancis.com/chapters/edit/10.1201/9781003540199-60/*
9. M. Kundu, B. Pasuluri and A. Sarkar, "Vehicle with Learning Capabilities: A Study on Advancement in Urban Intelligent Transport Systems," *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India, 2023, pp. 01-07, doi: 10.1109/ICAECT57570.2023.10118021.
10. K. Shekokar and S. Dour, "Epileptic Seizure Detection based on LSTM Model using Noisy EEG Signals," *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2021, pp. 292-296, doi: 10.1109/ICECA52323.2021.9675941.
11. S. J. Patel, S. D. Degadwala and K. S. Shekokar, "A survey on multi light source shadow detection techniques," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8275984.
12. K. Shekokar and S. Dour, "Identification of Epileptic Seizures using CNN on Noisy EEG Signals," *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, 2022, pp. 185-188, doi: 10.1109/ICECA55336.2022.10009127
13. A. Mahajan, J. Patel, M. Parmar, G. L. Abrantes Joao, K. Shekokar and S. Degadwala, "3-Layer LSTM Model for Detection of Epileptic Seizures," *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 2020, pp. 447-450, doi: 10.1109/PDGC50313.2020.9315833
14. P. Gin, A. Shrivastava, K. Mustal Bhihara, R. Dilip, and R. Manohar Paddar, "Underwater Motion Tracking and Monitoring Using Wireless Sensor Network and Machine Learning," *Materials Today: Proceedings*, vol. 8, no. 6, pp. 3121–3166, 2022
15. S. Gupta, S. V. M. Seeswami, K. Chauhan, B. Shin, and R. Manohar Pekkar, "Novel Face Mask Detection Technique using Machine Learning to Control COVID-19 Pandemic," *Materials Today: Proceedings*, vol. 86, pp. 3714–3718, 2023.
16. K. Kumar, A. Kaur, K. R. Ramkumar, V. Moyal, and Y. Kumar, "A Design of Power-Efficient AES Algorithm on Artix-7 FPGA for Green Communication," *Proc. International Conference on Technological Advancements and Innovations (ICTAI)*, 2021, pp. 561–564.
17. S. Chokoborty, Y. D. Bordo, A. S. Nenoty, S. K. Jain, and M. L. Rinowo, "Smart Remote Solar Panel Cleaning Robot with Wireless Communication," *9th International Conference on Cyber and IT Service Management (CITSM)*, 2021

10.48047/jocaaa.2024.33.05.35


18. V. N. Patti, A. Shrivastava, D. Verma, R. Chaturvedi, and S. V. Akram, "Smart Agricultural System Based on Machine Learning and IoT Algorithm," *Proc. International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2023.
19. S. T. Siddiqui, H. Khan, M. I. Alam, K. Upreti, S. Panwar and S. Hundekari, "A Systematic Review of the Future of Education in Perspective of Block Chain," in Journal of Mobile Multimedia, vol. 19, no. 5, pp. 1221-1254, September 2023, doi: 10.13052/jmm1550-4646.1955.
20. P. William, G. Sharma, K. Kapil, P. Srivastava, A. Shrivastava and R. Kumar, "Automation Techniques Using AI Based Cloud Computing and Blockchain for Business Management," *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, 2023, pp. 1-6, doi:10.1109/ICCAKM58659.2023.10449534.
21. A. Rana, A. Reddy, A. Shrivastava, D. Verma, M. S. Ansari and D. Singh, "Secure and Smart Healthcare System using IoT and Deep Learning Models," *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 2022, pp. 915-922, doi: 10.1109/ICTACS56270.2022.9988676.
22. Neha Sharma, Mukesh Soni, Sumit Kumar, Rajeev Kumar, Anurag Shrivastava, Supervised Machine Learning Method for Ontology-based Financial Decisions in the Stock Market, ACM Transactions on Asian and Low-Resource Language InformationProcessing, Volume 22, Issue 5, Article No.: 139, Pages 1 – 24, https://doi.org/10.1145/3554733
23. Sandeep Gupta, S.V.N. Sreenivasu, Kuldeep Chouhan, Anurag Shrivastava, Bharti Sahu, Ravindra Manohar Potdar, Novel Face Mask Detection Technique using Machine Learning to control COVID'19 pandemic, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3714-3718, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.368.
24. Shrivastava, A., Haripriya, D., Borole, Y.D. *et al.* High-performance FPGA based secured hardware model for IoT devices. *Int J Syst Assur Eng Manag* 13 (Suppl 1), 736–741 (2022). https://doi.org/10.1007/s13198-021-01605-x
25. A. Banik, J. Ranga, A. Shrivastava, S. R. Kabat, A. V. G. A. Marthanda and S. Hemavathi, "Novel Energy-Efficient Hybrid Green Energy Scheme for Future Sustainability," *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, Uzbekistan, 2021, pp. 428-433, doi: 10.1109/ICTAI53825.2021.9673391.
26. K. Chouhan, A. Singh, A. Shrivastava, S. Agrawal, B. D. Shukla and P. S. Tomar, "Structural Support Vector Machine for Speech Recognition Classification with CNN Approach," *2021 9th International Conference on Cyber and IT Service Management (CITSM)*, Bengkulu, Indonesia, 2021, pp. 1-7, doi: 10.1109/CITSM52892.2021.9588918.
27. Pratik Gite, Anurag Shrivastava, K. Murali Krishna, G.H. Kusumadevi, R. Dilip, Ravindra Manohar Potdar, Under water motion tracking and monitoring using wireless sensor network and Machine learning, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3511-3516, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.283.
28. A. Suresh Kumar, S. Jerald Nirmal Kumar, Subhash Chandra Gupta, Anurag Shrivastava, Keshav Kumar, Rituraj Jain, IoT Communication for Grid-Tie Matrix Converter with Power Factor Control Using the Adaptive Fuzzy Sliding (AFS) Method, Scientific Programming, Volume, 2022, Issue 1, Pages- 5649363, Hindawi, *https://doi.org/10.1155/2022/5649363*
29. A. K. Singh, A. Shrivastava and G. S. Tomar, "Design and Implementation of High Performance AHB Reconfigurable Arbiter for Onchip Bus Architecture," *2011 International Conference on Communication Systems and Network Technologies*, Katra, India, 2011, pp. 455-459, doi: 10.1109/CSNT.2011.99.