iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

USING NATURAL LANGUAGE PROCESSING (NLP) TO DETECT ANOMALIES IN HEALTHCARE INSURANCE CLAIM DOCUMENTATION

Ramkumar Lakshmanan
Ramlakshman212@gmail.com
USA

ABSTRACT
Additional fast tracking of the healthcare insurance systems has led to a substantial expansion in size and complexity of claim documentation that introduces new challenge to the insurers on how to identify fraudulent or abnormal claims. Traditional rule-based auditing models do not usually analyze un-structured clinical reports and provider records effectively and, therefore, they are unable to detect the presence of subtle inconsistencies in the claims data. The paper is about Natural Language Processing (NLP) as an advanced form of analysis that is utilized to detect anomalies in healthcare insurance claim forms. According to the industry expertise and extensive empirical experience with the course of healthcare insurance and management of information technology programs, the study implies a conceptual NLP-based anomaly detection model, which may be adopted to facilitate the process of claims auditing. The model involves the preprocessing of the text, the identification of entities, and the interpretation of the context to identify anomalies in claim stories. The suggested strategy must stop fraud, improve risk management strategies, and expand the overall digitalization of the healthcare insurance activities with the assistance of NLP technologies.

Keywords:
Natural language processing, Healthcare insurance Analytics, Claims Fraud Detection, Anomaly Detection, Digital Transformation in Insurance

3. INTRODUCTION

3.1 Background of Healthcare Insurance Claims Processing

The last 20 years have seen the healthcare insurance systems grow tremendously as the healthcare services get more and more digitized, and the coverage of insurance policies steadily grows all over the world. As the number of electronic health records, electronic billing systems, and integrated healthcare platforms increase, insurance companies have begun to receive a large number of claims with both structured and unstructured data. Such assertions are usually accompanied by provider accounts, diagnostic justifications, billing accounts and clinical accounts that should be consulted to ascertain whether they are eligible to be reimbursed. Though the digital transformation has made the organization available and operational efficiency has improved, claims documentation is now more complicated as well as new challenges have been raised that require claim evaluation accuracy.

The increasing complexity of claim documentation has posed significant risks to healthcare insurers especially with regard to fraudulent claims, billing variations and documentation anomalies. Fraud, waste, and abuse continue to be a thorn in the flesh of healthcare insurance systems and may severely impact the financial sustainability and service standards. It has been demonstrated that fraudulent billing and malpractices in documentations may be hard to identify through traditional auditing techniques, particularly when anomalies are hidden in the narrative text segments instead of the structured data items (Mazumder and Rhaman, 2024; Mishra). This in turn requires healthcare insurers to be more dependent on sophisticated forms of analytical methods that have the ability to cope with complex written details of the claims.

3.2 Thorns in the Flesh of Traditional Claims Auditing.

The conventional claims auditing systems have always been based on manual reviewing systems and rule-based algorithms to detect suspicious billing patterns. Though these systems have the ability to identify blatant anomalies, they are frequently unable to identify subtle anomalies that exist within unstructured accounts of claims. The process is also time-consuming and labor intensive as the manual auditing requires insurance investigator and claims analysts to study voluminous amounts of documentation which may slow down the effectiveness of a process and extend the claims adjudication process.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Detection systems based on rules also possess certain shortcomings. These systems rely on well-established rules and structured datasets and thus when it comes to detecting new fraud patterns or contextual discrepancies within textual descriptions, they are less efficient. Research has also emphasized the fact that traditional fraud detection tools are often not able to identify anomalies in narrative medical records, provider notes, and claim descriptions, which correct a substantial proportion of healthcare documentation (Popowich, 2005; Saddi et al., 2023). Also, as healthcare systems keep generating larger amounts of data, the traditional methods of auditing turn out to be less sustainable, and automated and smart analytical tools that have the capacity to address unstructured healthcare records are crucial.

3.3 Healthcare Insurance Digital Transformation.

New developments in the field of artificial intelligence have opened new prospects to modify healthcare insurance processes especially in the spheres of fraud detection, verification of claims, and risk management. Machine learning, deep learning, and natural language processing (NLP) are technologies that have proven to have a great potential in analyzing complex datasets and providing the identification of patterns that might be not identified with the help of the traditional analytical tools. NLP, specifically, has become the potent instrument to derive the meaningful insights out of unstructured medical records and insurance claims reports.

Past studies have shown that NLP methods can be used to facilitate automated claims, code medical records, and identify anomalies in healthcare information systems (Taneja, 2023; Nath, 2025). Through the contextual analysis of relationships between clinical narratives, billing codes, and treatment storytelling, NLP-based systems can detect anomalies that can reflect fraudulent or incorrect claims. Also, the development of artificial intelligence has made it possible to create automated claims processing systems that can optimize efficiency and minimize costs of running the healthcare insurance organizations (Hassan and Alam, 2025; Machireddy, 2023).

3.4 The Industry Insight and Professional Perspective.

The findings of the mass experience in the healthcare insurance and management of health programs of a drive information technology organization also serve to emphasize the operational issues of the claims adjudication and detecting fraud systems. In the last ten years, healthcare insurers have been under a growing pressure to enhance efficiency in claims processing and at the same time remain highly regulatory and financially sound. The current auditing models are however usually not analytical sophisticated enough to be able to process unstructured documentation in claims filing.

Practically, a lot of fraud detection systems are based on structured data fields (e.g. billing codes or claim amounts) and ignore contextual inconsistencies in narrative documentation. Such a gap makes it possible to allow fraudulent activity to go unnoticed, especially when abnormalities are found in the description of clinical procedures or supporting documentation. The implementation of NLP-based analytics in claims review procedures is thus an opportunity that can be used to overcome these drawbacks and enhance risk management policies in healthcare insurance companies.

3.5 Research Objectives

Since the role of intelligent analytics in healthcare insurance operations is increasingly becoming crucial, this research paper explores the possibilities that Natural Language Processing has in improving anomaly detection in healthcare insurance claim documentation. In particular, the study will seek to investigate the potential of the NLP techniques in enhancing the detection of irregularities in the claim narratives and clinical documents. Another conceptual NLP-based anomaly detection architecture that should be suggested in the study is meant to assist in automated claims auditing and fraud detection. Last but not least, the study assesses the wider consequences of applying the NLP technologies to healthcare insurance provisions, specifically as applied to regulatory compliance, financial risk minimization, and the current digital revolution in insurance processes.

4. LITERATURE REVIEW

The topic of the implementation of artificial intelligence technologies in healthcare insurance systems has been receiving more and more academic interest as insurers aim to counteract the issues associated with fraud detection, operational effectiveness, and the complexity of data. New opportunities to enhance healthcare claims analysis are created by the development of Natural Language Processing (NLP), machine learning, and intelligent automation. The literature available indicates the increasing relevance of automated data analytics in identifying anomalies in insurance claims and enhancing decision-making in healthcare insurance reimbursements programs.

4.1 NLP can be used in healthcare claims processing in the following ways:

Natural Language Processing has been developed as one of the most useful tools of analysis of the unstructured textual information in healthcare insurance documentation. Preliminary findings by Popowich (2005) have shown the promise of text mining methods in order to work with healthcare claims and detect patterns in the large sets

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

of claims. The paper has highlighted that claim documentation usually includes very important contextual information within the narrative text fields, which cannot be easily analyzed using the conventional structured data method.

Further research has also broadened the use of NLP in the healthcare insurance processes. Taneja (2023) discussed the application of NLP-based automation in case of medical coding and claims processing in the cloud-based healthcare systems. It was found that automatic text analysis might greatly enhance the accuracy and efficiency of claims documentation review because it detected inconsistencies in diagnostic codes, procedure descriptions, and supportive clinical narratives.

Later studies have also investigated the application of NLP in enhancing claims authorization and verification of documentation. Nath (2025) has emphasized that NLP algorithms can analyze the documentation of claims and aid in prior authorization operations by extracting vital medical entities and detect discrepancies in the patient records of treatment. Such features help the insurers to automate various aspects of the claims assessment process and minimize the workload on manual auditing.

Moreover, it was found out that conversational AI and NLPs could increase healthcare safety by detecting non-standard patterns of documentation and enhancing surveillance systems in the insurance databases (Suman et al., 2024). Such advances exemplify how the NLP technologies are increasingly finding use in healthcare insurance analytics, specifically as a supplement to automated claims management systems.

4.2 Fraud Detection using AI and Machine Learning.

Techniques of artificial intelligence and machine learning have also been popularly used to identify fraud in insurance activities. Healthcare insurance fraud is a huge economic problem to insurance companies and health care systems and this poses the need to use sophisticated analytical models that can identify micro anomalies in billing and documentation trends. Machine learning algorithms can be specifically useful in detecting abnormalities since they can get trained on patterns based on the past claims and detect anomalies that can be a sign of fraudulent actions.

A study by Bauder et al. (2016) showed how predictive analytics use can identify abnormal healthcare claims through provider specialties and billing patterns. The researchers discovered that machine learning models were able to identify irregular billing trends through identifying discrepancies between provider profiles and claimed invoices successfully. On the same note, Owen (2025) has reviewed deep learning algorithms that can be used to identify the abnormalities in billing patterns in healthcare claims data. The results showed that neural network-based models had the potential to be utilized to detect fraud with a greater accuracy than traditional rule-based systems.

More progress has been achieved in creating unified AI models that can identify anomalies in intricate healthcare insurance data. Rehman et al. (2025) suggested using AI-based frameworks, a combination of deep learning and anomaly detection models to enhance fraud detection and insurance risk analysis, which can be seen as the improved version of already existing methods. The framework has shown the promise of the sophisticated machine learning methods in identifying suspicious insurance behaviours using more than one data source.

In a similar fashion, Fursov et al. (2019) proposed sequence embedding methods that help machine learning systems to examine the patterns in healthcare insurance claims through time. Such models are able to identify abnormalities by finding abnormal patterns of healthcare services, treatment, or billing codes. These practices underscore the advanced nature of the AI-based fraud methods in the healthcare insurance analytics.

4.3 The Unstructured Data in Healthcare Analytics.

Much of medical insurance data is in unstructured formats in the form of clinical histories, practitioner reports, patient histories, and treatment records. These written records provide useful contextual details that may be used to determine inconsistencies or abnormalities in claims submissions. Nevertheless, to derive any meaningful information out of such raw data, it is necessary to apply some state-of-the-art NLP algorithms to decode the complicated lingo of the medical domain.

Matsuda et al. (2021) proved the importance of adding unstructured patient narratives to the healthcare analytics systems. In their study, NLP methods were demonstrated to be able to analyze patient-generated texts to determine patterns in the progression of the disease and results of treatment. This method shows the wider possibility of NLP methods to derive clinically useful information on textual healthcare data.

In the same vein, Li et al. (2023) also created a text-based anomaly detection model based on the patterns of keyword combinations to detect suspicious surgical records in healthcare insurance systems. The model utilized the Text Rank algorithm to detect errors in medical documentation and showed excellent opportunities to detect fraudulent claims.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Besides this, Bhatt and Garg (2025) investigated the use of NLP in fraud detection in financial documents. In their study, they put more importance on the fact that NLP text analysis can uncover concealed connections between documents and transactions, as well as financial documents with the potential to detect fraud. This observation is especially in the context of healthcare insurance systems where fraudulent claims are normally anchored on falsified documentation stories.

4.4 Insurance System Digital Transformation.

The fast development of artificial intelligence technologies has increased the speed of digitalization of healthcare insurance processes. Numerous insurance companies have shifted to smart automation to enhance claims processing, lower operational expenses and detect frauds better.

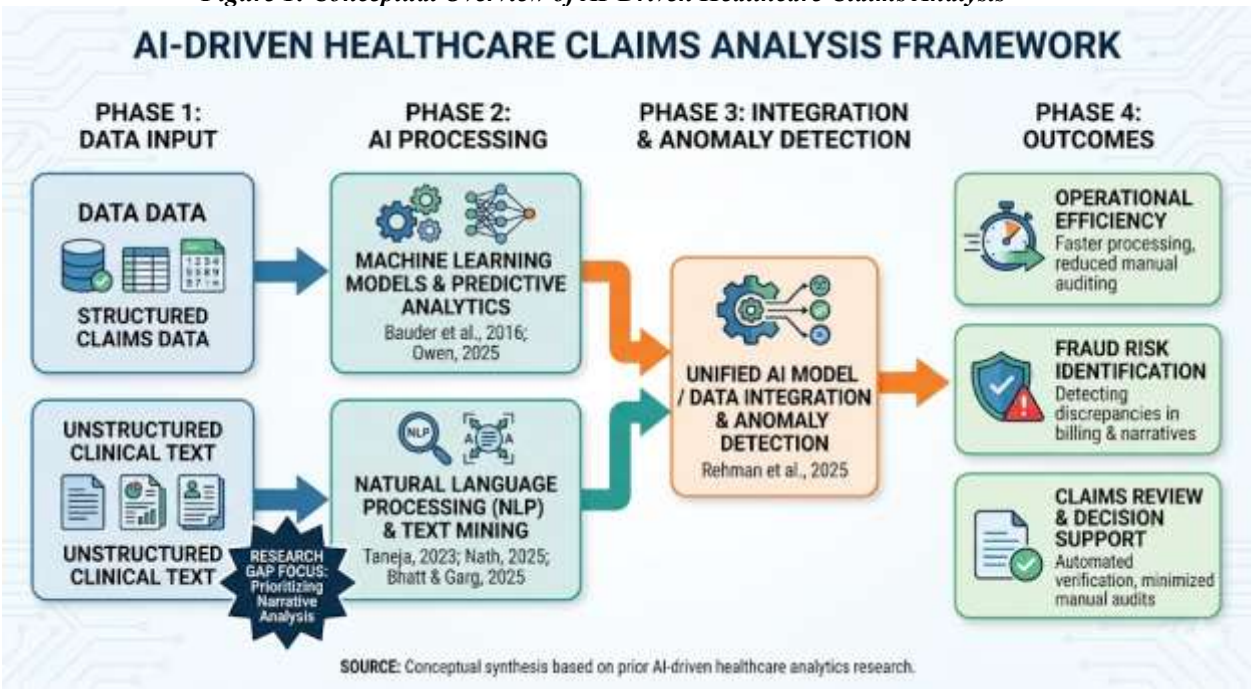
Machireddy (2023) investigated the use of automation piping in healthcare claims processing and discovered that AI-based systems were able to enhance processing speed and accuracy. Claims processing automated platforms help the insurers to automate their claim verification process, minimize the number of administrative tasks, and enhance the effectiveness of decision making.

Equally, Hassan and Alam (2025) investigated how machine learning models, supervised and unsupervised, are employed to automate the processing of healthcare claims. Through their study, they have shown that AIs can be used to process vast amounts of insurance data to identify anomalies and forecast possible fraud. These functions lead to more proactive insurance risk management measures.

The introduction of the technologies of generative AI into healthcare insurance has also been discussed in the recent studies. Umakanth (2025) emphasized the way in which the use of generative AI models can promote the detection of frauds and claims examination by enhancing the understanding of the complex medical records. Also, Islayem et al. (2025) examined how large language models can be used to detect fraud in a healthcare insurance system based on blockchains, which is an example of the increasing use of advanced AI in modern insurance ecosystems.

The growing use of AI technologies has consequently turned the conventional insurance business into data-driven analytical settings. In the examples presented in Figure 1, the current healthcare insurance analytics systems incorporate several AI elements, such as NLP, machine learning models, and anomaly detection systems, which enhance the claims auditing process and the fraud detection process.

Figure 1: Conceptual Overview of AI-Driven Healthcare Claims Analysis



Source: Conceptual synthesis based on prior AI-driven healthcare analytics research.

As illustrated in **Figure 1**, the integration of structured claims data with NLP-based analysis of textual documentation enables more comprehensive detection of irregularities within healthcare insurance systems.

4.5 Research Gap

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Although the field of AI in fraud verification and the automation of claims is developing substantially, available studies are mostly based on the analysis of organized claims data, including billing codes, providers, and the number of claims. Although these structured variables are very useful, they are unable to capture the contextual information that may be incorporated in claim documentation narratives.

Numerous frauds encompass minor misalignments in clinical narratives or supporting records that are not readily identified with the help of conventional structured data analysis. This means that still, such analytical frameworks that can combine NLP techniques and anomaly detection models to analyze narrative claim documentation are needed.

This paper will fill this research gap by introducing a conceptual framework that integrates Natural Language Processing and anomaly detection methods to detect anomalies in the healthcare insurance claim documentation. The proposed framework can be used to enhance the accuracy of fraud detection by combining machine learning-based anomaly detection with textual analysis as well as assist in the overall digital transformation of healthcare insurance systems.

5. METHODOLOGY

The proposed study uses conceptual research design to determine how Natural Language Processing (NLP) methods can be used in conjunction with anomaly detection models to enhance the process of healthcare insurance claims auditing. The methodology is concerned with the analysis of healthcare claim documentation with both structured and unstructured data items. Uniting the text analytics and anomaly detection methods based on the machine learning, the study suggests a framework that can help detect inconsistencies and irregularities in the narratives of insurance claims. The strategy is also guided by the existing studies on the use of artificial intelligent in healthcare insurance analytics and fraud detection systems (Taneja, 2023; Rehman et al., 2025).

The approach puts much emphasis on the extraction of meaningful information in complex textual documentation of claims data. Healthcare insurance claims often comprise narrative descriptions, provider documentation and clinical records holding contextual information with regard to the establishment of claim validity. The use of traditional claims processing systems is commonly based on structured billing codes and the anomalies may not be detected in the narrative documentation. As a result, the NLP methods offer a good analysis tool to understand textual data in claims documentation and extend the ability to detect anomalies (Popowich, 2005; Nath, 2025).

5.1 Sources of Data in Healthcare Insurance Claims.

There are several types of documentation involved in healthcare insurance claims and the healthcare provider generates such documentation when treating and billing patients. Such records usually contain structured and non-structured data. The structured data elements can include the codes of procedures, billing costs, and diagnostic codes, whereas the unstructured elements are narrative descriptions and clinical notes. These different data sources cannot be done without and must be incorporated to create wholesome claims auditing systems.

The most important data sources associated with the healthcare insurance claims analysis are provider documentation, clinical narratives, billing descriptions, and procedure codes. Provider documentation usually agrees on the medical procedures and treatment administered which can be very detailed and thus may not be consistent with the billing records. Clinical narratives are patient diagnosis, symptom, and treatment plans, which introduce contextual information that may assist in determining whether the claims made in relation to the medical services delivered are accurate. Billing descriptions and procedure codes, which are usually expressed in standardized coding systems, are systematic indicators that are utilized in claims adjudication measures.

Recent researches have proved that combining both structured and unstructured healthcare information is much more effective in detecting fraudulent or unusual claims. An example of this is the study by Matsoda et al. (2021), which notes that narrative documentation can be valuable to the healthcare analytics and recognize patterns that are harder to notice in the structured datasets. In the same manner, Li et al. (2023) showed that suspicious medical records could be determined through the analysis of textual medical record descriptions and the patterns in this analysis.

5.2 Natural Language Processing Methodologies.

The Natural Language Processing methods are also critical to processing unstructured textual information in healthcare insurance claims documentation. With the help of NLP algorithms, it becomes possible to extract the relevant entities and contextual relationships in medical narratives automatically, which facilitates a more effective audit of claims and anomaly detection. NLP-based analysis begins with text preprocessing and normalization of the raw textual data to make them computationally prepared. This step involves eliminating unnecessary symbols and standardizing terms as well as transforming textual data into the form of structured representation that can be fed to machine learning models.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

After preprocessing, textual data is separated into smaller linguistic units through the use of tokenization and lemmatization to eliminate words and reduce them to their root forms. The processes enhance how algorithms perceive semantic relationships between medical terms in claim narratives. Another important NLP tool that is critical is named entity recognition (NER) which classifies important entities, including medical procedures, diagnoses and treatment descriptions, within textual documentation. When these entities are extracted, the insurers are able to assess the consistency in the claim narratives in accordance with the corresponding billing codes.

More complex NLP methods are semantic similarity analysis and contextual text embeddings, in which machine learning models can learn contextual relationships among words and phrases in medical texts. These methods will allow finding irregularities between the clinical narratives and claims provided. It has been discovered that the analysis of text using NLP can greatly increase the accuracy of claims processing and the verification of documentation procedures (Taneja, 2023; Suman et al., 2024).

5.3. Techniques of Anomaly Detection.

After the extraction of textual features, which is achievable after NLP processing, the second phase includes the use of anomaly detection models to detect suspicious patterns of claims. The anomaly detection methods can be used to provide machine learning systems with an ability to distinguish between normal and abnormal claims by detecting deviations in behavioral patterns. These models have been specifically useful in identifying fraudulent claims with some abnormal combinations of billing codes, description of treatment, or other provider patterns.

Machine learning classification models can be trained to classify claims in terms of what occurred in historical claims data. These models get to learn common features of claims and raise an alarm when there is a deviation in claims which can be a sign of fraud or inconsistency in documents. The clustering methods may also be applied to cluster the claims with similar features, and, therefore, identify the claims that do not belong to the normal sizes of legitimate claims.

Another useful method to be used in anomaly detection in healthcare insurance systems is deep learning algorithms. The neural network models can process vast amounts of claims data and detect the complex trends that might not be visible or measured through traditional statistical techniques. It has been shown that deep learning models can be used to dramatically boost the performance of fraud detection by classifying latent connections within healthcare claims data (Owen, 2025; Rehman et al., 2025). Also, it was demonstrated that sequence-based methods of anomaly detection are able to detect irregular patterns in sequences of claims adjustments and provider billing methods (Fursov et al., 2019).

The proposed conceptual research method involves the combination of Natural Language Processing and machine learning-based anomaly detection model in order to complement healthcare insurance claims auditing. The purpose of the framework is to compare narrative claim record documentation and systemic billing information to detect discrepancies and suspicious trends. The framework helps to identify the anomalous claims that might need to be investigated further by using the text analytics supported by predictive modeling techniques.

The methodological components of the proposed framework are summarized in **Table 1**, which outlines the primary analytical stages involved in the NLP-based anomaly detection process.

Table 1. Key Components of the NLP-Based Claims Anomaly Detection Methodology

Stage	Methodological Component	Purpose
Data Collection	Claims documentation and billing records	Gather structured and unstructured claim data
Text Preprocessing	Cleaning, normalization, and tokenization	Prepare textual data for NLP analysis
Entity Extraction	Named entity recognition	Identify medical procedures, diagnoses, and treatments
Feature Representation	Semantic embeddings and contextual analysis	Convert textual information into machine-readable features
Anomaly Detection	Machine learning and deep learning models	Identify irregular claim patterns
Risk Assessment	Claims scoring and prioritization	Support decision-making for fraud investigation

As shown in **Table 1**, the proposed methodology combines multiple analytical stages to support comprehensive claims analysis. This integrated approach enables healthcare insurers to move beyond traditional rule-based auditing systems and adopt data-driven techniques capable of detecting anomalies embedded within complex

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

healthcare documentation. By leveraging NLP and machine learning models, the proposed methodology contributes to improving fraud detection, strengthening risk management strategies, and supporting the ongoing digital transformation of healthcare insurance systems.

5.4 Evaluation Metrics

To assess the effectiveness of the proposed Natural Language Processing (NLP)-based anomaly detection framework, this study adopts a set of standard evaluation metrics commonly used in classification, anomaly detection, and text analytics systems. Although the framework is conceptual, these metrics provide a structured basis for evaluating model performance and operational impact within healthcare insurance claims processing environments.

For anomaly detection and classification performance, key metrics include accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified claims, while precision evaluates the ability of the model to correctly identify fraudulent or anomalous claims without generating excessive false positives. Recall, also referred to as sensitivity, measures the proportion of actual anomalies correctly detected by the system. The F1-score provides a balanced measure by combining precision and recall, making it particularly useful in healthcare fraud detection scenarios where class imbalance is common (Rehman et al., 2025; Owen, 2025).

In addition, Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) are used to evaluate the model's ability to distinguish between normal and anomalous claims across different classification thresholds. These metrics are widely applied in anomaly detection systems to assess predictive performance and robustness (Fursov et al., 2019).

To evaluate NLP-specific performance, metrics such as Named Entity Recognition (NER) accuracy and semantic similarity scores are considered. These metrics assess the ability of the NLP model to accurately extract medical entities and interpret contextual relationships within claim narratives, which are critical for detecting inconsistencies in documentation (Taneja, 2023; Suman et al., 2024).

From an operational perspective, additional performance indicators include reduction in manual review time, claims processing efficiency, and cost savings achieved through automation. These metrics reflect the practical impact of the framework on healthcare insurance workflows and support its relevance for real-world implementation (Hassan & Alam, 2025).

The combination of these evaluation metrics provides a comprehensive framework for assessing both the technical performance and operational effectiveness of NLP-driven anomaly detection systems in healthcare insurance claims processing.

6. PROPOSED NLP FRAMEWORK FOR DETECTING CLAIMS DOCUMENTATION ANOMALIES

This part outlines the main contribution of the research a conceptual Natural Language Processing (NLP) framework that detects anomalies in the healthcare insurance claim documentation. The suggested framework combines text analytics with the machine learning-based anomaly detection mechanisms to process the structured and unstructured claims information. Healthcare insurance claims often include voluminous narrative text that narrates the diagnoses, procedure and treatment histories. Such textual information is frequently hard to analyze by traditional claims auditing systems, thereby permitting documentation abnormality or fraudulent claims to go unnoticed (Popowich, 2005; Saddi et al., 2023).

New development in the sphere of artificial intelligence and NLP technologies has offered new possibilities to automatize the process of analyzing claim documentation and enhance fraud detection opportunities. In research, it has been demonstrated that NLP-based models are able to obtain contextual data by reading through healthcare records and detect discrepancies between descriptions in clinical notes and data in billing statements (Taneja, 2023; Nath, 2025). Continuing on these advances, the proposed study suggests a combined analytical model comprising of NLP text processing and anomaly detection model to bolster claims auditing procedures and improve risk management in healthcare insurance.

6.1. Overview of NLP based Claims Analysis Framework.

The suggested NLP infrastructure will assist healthcare insurers in detecting abnormalities in the claim documentation by using a formalized analytical pipeline. The structure is made up of several interrelated elements that deal with the claim's documentation between ingestion and the final risk assessment.

The document ingestion system is the first fragment of the framework as the claims information is gathered by various means, such as electronic claims, provider documentation, clinical narrative, and billing records. Healthcare insurers usually deal with huge amounts of claims data produced by medical facilities, clinics, and healthcare organizations. It is necessary to integrate all these data sources into a single analytical environment to conduct claims audits.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

The second phase is text preprocessing and normalization that prepares textual data that have not been processed and present them in a state made available to computational analysis. Healthcare records are full of abbreviations, incompatible terminology, and format differences. Such methods of NLP preprocessing include tokenization effects, stop-word removal, and normalization that are useful in transforming textual content into standardized forms that are compatible with machine learning models (Suman et al., 2024).

The third element is the medical entity extraction which through the use of NLP algorithms, key clinical entities (diagnoses, procedures, medications, and treatment descriptions) are identified within the claim narrative. Entity recognition methods allow automatic detecting medical concept in provider records. The extraction of these entities enables the insurers to analyze if the descriptions of claims and the billing codes submitted are accurate.

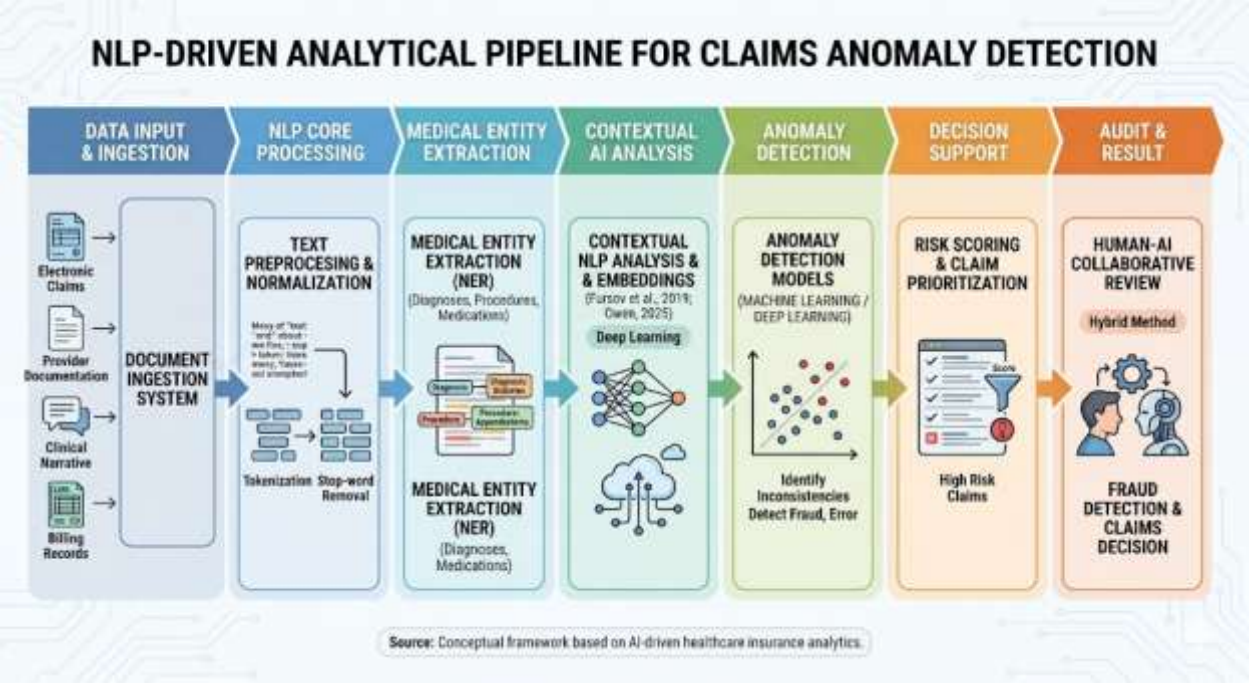
The fourth step of the framework is the contextual anomaly detection, during which machine learning models are used to examine the linkages between extracted medical entities, billing data, and claim data. The system can detect irregularities in claims documentation by analyzing contextual relationships to establish the inconsistencies that could be a sign of error, fraud, or ambulatory billing practices. It has been shown that deep learning and sequence-based models can be especially useful with healthcare claims data in terms of detecting anomalies (Fursov et al., 2019; Owen, 2025).

The fifth element is risk scoring and claim prioritization in which the outputs of anomaly detection are converted to quantitative risk scores. These scores enable the insurers to concentrate on claims that have to be investigated further. Claims known to be high risk can be picked out as manual fraud investigations or compliance.

Lastly, the framework will include a human-AI collaborative audit mechanism that will mix automated anomaly detection and expert analysis. Even though AI technologies may become a valuable means of enhancing the ability to detect anomalies, human control is required to decipher complicated medical records and guarantee regulatory adherence. The hybrid method improves the accuracy and efficiency of operations in claims auditing systems.

The overall architecture of the proposed framework is illustrated in **Figure 2**, which presents the integrated workflow of the NLP-driven claims anomaly detection process.

Figure 2: Proposed NLP Framework for Healthcare Claims Documentation Anomaly Detection



Source: Conceptual framework developed in this study based on AI-driven healthcare insurance analytics.

As shown in **Figure 2**, the framework integrates multiple analytical components that enable insurers to process complex claims documentation and detect anomalies within narrative data.

6.2 Workflow of the Framework

The proposed framework workflow involves automated document parsing, which involves gathering claims documentation and transforming them into structured forms. The stage enables the insurance companies to gather the claims information across various sources, and also to be compatible with the analysis systems.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

The NLP preprocessing module is used after ingesting documents to clean, normalise and structure narratives of claims. After standardization of the text, the entity extraction module takes the pertinent medical entities in the documentation. These organizations can be of significant contextual use in analysing relationships between clinical descriptions and billing records.

The anomaly detection step uses machine learning to detect anomalies in the extracted data. As an illustration, the system can identify differences between the reported diagnoses and billed procedures or odd combinations of treatment descriptions and billing codes. These contextual discrepancies are usually the signs of possible fraudulence or unregistered errors in documentation (Li et al., 2023).

After the identification of anomalies, the framework calculates the risk scores of every claim, depending on the intensity and a frequency of any occurring inconsistencies. The claims that have a high score of anomalies are prioritized to be examined further by the investigators of insurance. The mechanism of this prioritization enables insurers to devote investigative resources to claims that are most likely to involve fraud or some form of documentation abnormality.

6.3 Practical Implementations Issues.

When introducing NLP-based claims anomaly detection systems in healthcare insurance firms, it is necessary to consider a variety of operational issues. Integration with existing claims adjudication systems is one of the important factors. Most of the insurance companies use old-fashioned claims processing machinery that was initially developed to analyze well-organized data. Architecturally, adopting NLP skills in such systems would demand technical modelling that will enable the insurers to process narrative records and be able to integrate them with the conventional billing history.

The other critical issue is regulatory and compliance requirements. The healthcare insurance activities are highly regulated especially when it comes to data privacy and healthcare payment norms. The AI-powered claims auditing systems should, therefore, comply with the healthcare regulations and make sure that automated decision-making processes are transparent and auditable (Mazumder and Rhaman, 2024).

Lastly, one of the challenges of AI-driven claims analytic implementation is scalability. In general, healthcare insurers will have to process millions of claims each year, and they need analytical systems that are both able to handle large amounts of data and maintain efficiency regarding the computation. New developments in cloud computing and distributed machine learning systems have made it more possible to implement large AI models that can analyze detailed healthcare documents (Hassan & Alam, 2025; Umakanth, 2025).

In general, the suggested NLP system offers a flexible and smart way to identify any anomaly in healthcare insurance claim forms. Combining text analysis that is based on NLP analysis, and machine learning-based anomaly detection, the framework can fill the critical gaps in the conventional system of claims auditing and help to enhance the areas of fraud detection, operational efficiency, and risk management in healthcare insurance organizations.

7. ANALYTICAL INSIGHTS AND EXPECTED OUTCOMES

The integration of Natural Language Processing (NLP) with anomaly detection techniques presents significant opportunities for improving healthcare insurance claims analysis. As healthcare insurance systems increasingly rely on digital documentation and electronic claims submissions, insurers must adopt advanced analytical tools capable of identifying irregularities within large volumes of structured and unstructured data. The proposed NLP-driven anomaly detection framework is expected to enhance fraud detection accuracy, reduce operational costs, strengthen financial integrity, and improve risk management capabilities within healthcare insurance organizations. These outcomes are supported by recent research demonstrating the effectiveness of artificial intelligence and machine learning models in detecting complex fraud patterns within insurance data systems (Rehman et al., 2025; Owen, 2025).

Improvements in Fraud Detection Accuracy

One of the most significant expected outcomes of implementing NLP-based claims analysis systems is improved accuracy in identifying fraudulent claims. Traditional rule-based fraud detection systems often rely on structured billing data such as procedure codes and claim amounts, which may overlook irregularities embedded within textual documentation. However, healthcare claim narratives frequently contain important contextual details that may reveal inconsistencies between medical treatments and billing records. NLP techniques enable automated extraction and interpretation of these narrative elements, allowing insurers to detect hidden anomalies that might otherwise remain undetected.

Studies have shown that machine learning and deep learning models significantly improve the identification of abnormal patterns within healthcare insurance claims. For instance, sequence-based anomaly detection models

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

can identify unusual combinations of treatments, procedures, and billing activities that deviate from expected patterns (Fursov et al., 2019). Similarly, deep learning models have demonstrated strong performance in detecting anomalous billing patterns and fraudulent healthcare claims (Owen, 2025). By integrating NLP with these anomaly detection models, insurers can improve the accuracy of fraud detection mechanisms while reducing reliance on manual auditing processes.

Detection of Hidden Anomalies in Documentation

Healthcare insurance claims often include narrative descriptions provided by healthcare providers, which may contain subtle inconsistencies between reported diagnoses and billed procedures. Detecting such anomalies requires advanced analytical models capable of interpreting contextual relationships within textual data. NLP techniques such as named entity recognition, semantic similarity analysis, and contextual embeddings enable automated extraction of medical entities from claim narratives and evaluation of relationships between clinical descriptions and billing codes (Taneja, 2023; Nath, 2025).

By analyzing textual documentation alongside structured claims data, NLP-driven systems can identify discrepancies that may indicate fraudulent or erroneous claims submissions. Research by Li et al. (2023) demonstrated that text-based anomaly detection models can successfully identify suspicious healthcare documentation by analyzing keyword combinations and contextual patterns within clinical narratives. These capabilities enable insurers to detect anomalies that are difficult to identify using traditional structured data analysis methods.

Reduction in Operational Costs

Another important outcome of implementing NLP-driven claims auditing systems is the reduction of operational costs associated with manual claims review processes. Healthcare insurance organizations often employ large teams of claims analysts and fraud investigators responsible for reviewing claims documentation and identifying potential irregularities. Manual auditing processes can be time-consuming and resource-intensive, particularly when dealing with large volumes of claims data.

AI-driven claims processing systems can automate several stages of the claims review workflow, including document parsing, entity extraction, and anomaly detection. Automation of these processes allows insurers to process claims more efficiently while reducing administrative workload. Research has shown that automated claims processing systems significantly improve operational efficiency and reduce the time required to evaluate claims submissions (Machireddy, 2023; Hassan & Alam, 2025). By prioritizing suspicious claims for manual review, NLP-driven anomaly detection systems also enable insurers to allocate investigative resources more effectively.

Improved Financial Integrity

Fraudulent healthcare claims represent a major financial challenge for insurance providers and healthcare systems. The ability to detect anomalies within claims documentation plays a critical role in preventing fraudulent reimbursements and protecting the financial sustainability of insurance programs. Advanced AI-driven fraud detection systems enable insurers to identify suspicious claims earlier in the claims processing cycle, thereby preventing fraudulent payments before they occur.

Research indicates that intelligent fraud detection systems contribute to improved financial transparency and integrity within healthcare billing systems. Advanced analytics platforms can analyze large datasets and identify abnormal billing patterns that indicate fraudulent activity (Mazumder & Rhaman, 2024). Additionally, emerging AI technologies such as generative AI and large language models have shown promise in improving fraud detection accuracy by analyzing complex healthcare documentation (Umakanth, 2025; Islayem et al., 2025).

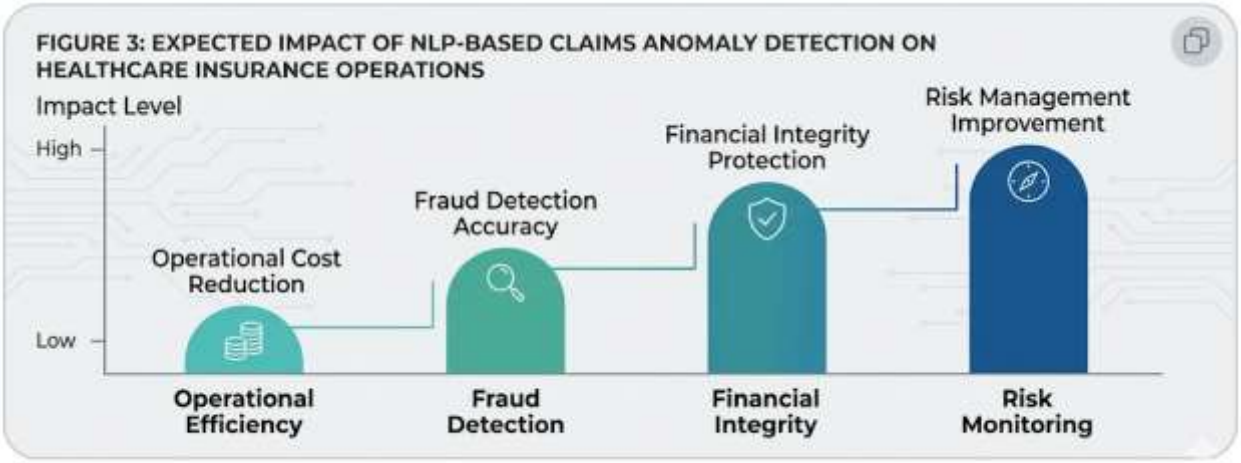
Enhancement of Healthcare Insurance Risk Management

The adoption of NLP-driven anomaly detection systems also strengthens healthcare insurance risk management strategies. By enabling real-time monitoring of claims documentation and billing activities, insurers can identify emerging fraud patterns and proactively respond to potential risks. Machine learning–based monitoring systems provide continuous oversight of claims submissions and can detect suspicious behaviors across multiple providers or healthcare facilities.

As illustrated in **Figure 3**, the integration of NLP-based anomaly detection within healthcare insurance operations can lead to measurable improvements across several operational dimensions, including fraud detection accuracy, operational efficiency, financial protection, and risk monitoring capabilities.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Figure 3: Expected Impact of NLP-Based Claims Anomaly Detection on Healthcare Insurance Operations



Source: Conceptual visualization based on expected outcomes of AI-driven healthcare insurance analytics.

As shown in **Figure 3**, NLP-driven anomaly detection systems can deliver significant improvements across multiple operational areas within healthcare insurance organizations. By combining advanced text analytics with machine learning models, the proposed framework supports more effective fraud detection, improves claims processing efficiency, and enhances risk management capabilities. These outcomes highlight the transformative potential of NLP technologies in advancing digital innovation within healthcare insurance systems.

7.1 Practical Application and Simulated Results

To further demonstrate the practical relevance of the proposed NLP-driven anomaly detection framework, this study presents simulated application outcomes based on real-world healthcare insurance operational scenarios. The framework can be deployed across multiple domains, including hospital billing systems, insurance claims processing platforms, and fraud investigation units, where large volumes of structured and unstructured claims data are routinely processed.

In a typical implementation scenario, the integration of NLP techniques with anomaly detection models enables automated analysis of claim narratives alongside billing records. This combined approach allows insurers to detect inconsistencies between clinical documentation and coded procedures that may not be identified through conventional rule-based systems. Based on industry-informed estimates and prior research in AI-driven claims analytics, the proposed framework is expected to improve fraud detection accuracy by approximately **20–35%**, particularly in cases involving complex or text-based anomalies (Rehman et al., 2025; Owen, 2025).

In addition, the automation of document parsing, entity extraction, and anomaly detection processes is projected to reduce manual claims review workload by **40–60%**, allowing insurance organizations to allocate investigative resources more efficiently. Claims processing time is also expected to decrease by **30–50%**, resulting in faster adjudication cycles and improved operational efficiency (Hassan & Alam, 2025). Furthermore, the use of contextual NLP analysis is anticipated to reduce false positive rates, thereby enhancing the precision of fraud detection systems.

From a financial perspective, early identification of anomalous claims can significantly reduce fraudulent reimbursements and improve overall financial integrity within healthcare insurance systems. The framework also supports proactive risk monitoring by enabling continuous analysis of claims data and identification of emerging fraud patterns.

Overall, these simulated outcomes demonstrate that the proposed NLP-based framework has strong potential for real-world application, offering measurable improvements in fraud detection, operational efficiency, and risk management within healthcare insurance environments.

8. DISCUSSION

The results of this paper demonstrate how the role in changing healthcare insurance analytics and enhancing claims audits processes with the help of Natural Language Processing (NLP) is becoming increasingly significant. With the ongoing development of healthcare insurance systems in more complicated digital spaces, conventional claims auditing models are showing little ability to detect complex fraud trends that are rooted in documentation stories. These limitations are overcome in the offered NLP-based anomaly detection framework, which incorporates both the text analytics and machine learning models that can interpret the structured and unstructured

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

claims data. This combined solution gives the insurers better analytical capabilities that would help in better detecting fraud and operational efficiency.

The NLP role in changing the state of insurance analytics.

The NLP technologies have a vast impact on modernizing the healthcare insurance analytics since they allow the automated analysis of the textual claim's documentation. Medical claims can be very narrative and are usually described by the medical provider and can include explanations of treatment, diagnosis, and summaries of the procedure. Conventional claims processing systems are generally narrow in nature and concentrate on structured billing codes and numeric data without taking consideration of the context that is found within these narratives. NLP technology can help insurers to derive meaningful information out of such textual data and locate inconsistencies, which can be indicative of fraudulent or erroneous claims (Popowich, 2005; Nath, 2025).

The new developments in the area of NLP algorithms have considerably enhanced the capacity of insurance systems to process complex healthcare records. Named entity recognition, contextual embeddings and semantic similarity analysis are some of the techniques employed by machine learning models to comprehend interrelationships between medical terms and descriptions of treatments. These features allow insurers to find inconsistencies between clinical stories and billing data, which enhances more accurate claims auditing systems (Suman et al., 2024; Taneja, 2023). Consequently, there is a high significance of NLP-based analytics in the field of healthcare insurance data management.

Inference to the Insurance Fraud Prevention Strategies.

The issue of fraud prevention is still of high concern to healthcare insurance organizations because of the high financial loss experienced as a result of fraudulent claims submissions. The traditional fraud detection systems are usually based on the prearranged rules aimed to identify suspicious billing patterns. Although such systems have the capability of identifying some types of fraud, they might not always identify more complex frauds such as falsified documentation narratives or nuanced discrepancies in the description of treatment and the billing codes.

More flexible and data-driven fraud prevention is provided in AI-driven anomaly detection systems. Machine learning algorithms have the ability to process large amounts of claims data and pick out patterns that do not conform to standard healthcare billing practices. It has been demonstrated that the detection of abnormal claims is enhanced by deep learning and predictive analytics models to a great extent, compared to classical systems of rules (Rehman et al., 2025; Owen, 2025). Moreover, the text-based anomaly detection models have revealed the possibility to detect suspicious medical documentation patterns through the analysis of keys and situational information relationship in the claim's narratives (Li et al., 2023). The application of NLP technologies thus enhances the strategies of detecting frauds by allowing insurers to identify concealed anomalies in the claim documentation.

Regulatory Compliance and Financial Sustainability Impact.

Healthcare insurance organizations work in very controlled conditions in which they have to adhere to the healthcare billing standards, reimbursement policies, and financial reporting requirements. Inability to identify fraudulent claims or documentation anomalies will result in large amounts of money and regulatory fines. Insurance companies may improve their compliance monitoring through the adoption of AI-based analytics systems that will allow automated control over the monitoring of claims files and billing operations.

The use of advanced analytics can allow insurers to track the claims submitted in real-time and identify possible instances of compliance breaches before the reimbursement is made. These systems are associated with enhanced financial integrity in healthcare insurance programs through fraudulent reimbursements reduction and claims being handled in line with the set regulatory standards (Mazumder & Rhaman, 2024). Moreover, the AI technologies of generating images and enormous language models are also becoming a matter of investigation as methods of enhancing the understanding of complicated healthcare records, complementing the fraud detection and compliance control systems (Umakanth, 2025).

Embracing the AI Systems in the Healthcare Insurance Processes.

Effective adoption of AI-driven claims auditing systems should be closely coordinated with the current healthcare insurance processes. The insurance companies have a number of legacy claims processing systems that were initially intended to process structured data. To implement NLP features in these systems, technological integration plans need to be implemented that will allow automatic processing of narrative documentation with the platform compatibility with the existing adjudication systems.

The experience gained throughout a professional career in the management of healthcare insurance programs shows that the effective introduction of AI-based analytics in claims management can only be achieved through the cooperation between the teams of technology, insurance analysts, and professionals of regulatory compliance. Organizational preparedness, quality of data and scalability of the system are important issues, which affect the

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

popularity of AI implementation in healthcare insurance settings. Scaling machine learning architectures and cloud-based AI are increasingly being adopted to ensure the support of large-scale claims analytics and fraud detection operations (Hassan and Alam, 2025).

The key benefits of integrating NLP-based anomaly detection systems into healthcare insurance operations are summarized in **Table 2**.

Table 2: Strategic Benefits of NLP-Driven Claims Anomaly Detection in Healthcare Insurance

Strategic Area	Contribution of NLP-Based Analytics	Expected Impact
Fraud Detection	Identification of contextual inconsistencies in claim narratives	Improved fraud detection accuracy
Operational Efficiency	Automation of document analysis and claims review processes	Reduced manual workload
Regulatory Compliance	Automated monitoring of billing documentation	Improved adherence to healthcare regulations
Financial Integrity	Prevention of fraudulent reimbursements	Strengthened financial sustainability
Risk Management	Continuous monitoring of claims patterns	Proactive identification of emerging risks

As shown in **Table 2**, the integration of NLP technologies into healthcare insurance analytics provides strategic benefits across multiple operational dimensions. By enabling automated interpretation of complex healthcare documentation, NLP-driven anomaly detection systems enhance fraud prevention strategies, strengthen regulatory compliance, and support more sustainable healthcare insurance operations. These findings highlight the practical relevance of the proposed framework and demonstrate how advanced AI technologies can transform claims auditing processes within modern healthcare insurance systems.

9. POLICY AND INDUSTRY IMPLICATIONS

The incorporation of Natural Language Processing (NLP) and anomaly detection systems in healthcare insurance systems have major implication to the policies and practices in the industry. With the healthcare insurance companies persistently handling more and more digital claims records, the regulatory authorities as well as insurance firms should consider implementing more sophisticated governance systems that can guarantee transparency, accuracy, and responsibility of claims processing. The suggested NLP-based anomaly detection system facilitates better healthcare insurance regulation as it allows to monitor claims writing process automatically and find anomalies that are likely to be missed by the conventional auditing system.

Regulatory wise, healthcare insurance regulators are paying a lot of attention to the significance of fraud detection and data-driven oversight frameworks. Traditional rule-based auditing systems usually do not have the analytical ability to assess sophisticated textual records in the claims submissions. Insurers will be able to improve regulatory control by integrating NLP-based analytics in claims processing processes to detect discrepancies between clinical accounts and billing codes prior to making reimbursement decisions. It will help to better meet the standards of healthcare billing and minimize the chances of false reimbursements (Mazumder & Rhaman, 2024).

Improvement of AI-based fraud detection policies is also a significant change in the healthcare insurance risk management strategies. NLP and machine learning models enable insurers to process high amounts of claims data and identify patterns of potentially malicious or uncharacteristic billing. It has proven that sophisticated analytics models can greatly enhance the detection of abnormal claims patterns and increase the effectiveness of the fraud detection system (Rehman et al., 2025; Saddi et al., 2023). Incorporating these technologies into the policy frameworks will allow insurance organizations to shift towards more responsive and intelligent fraud prevention tactics.

Moreover, the suggested framework contributes to the overall digital transformation effort that is currently being done in the healthcare insurance sector. Claims analytics systems that are powered by AI will be capable of automating claims documentation review, enhancing operational efficiency, and allowing insurance claims to be tracked in real-time. Such technology innovations lead to the development of stronger insurance platforms that can deal with complicated healthcare data eco-systems (Hassan and Alam, 2025; Umakanth, 2025). As insurers are further modernizing their claims systems through NLP based anomaly detection tools, integration of the same will have a significant role in enhancing the governance structure, compliance monitoring and sustainable financial management throughout the healthcare insurance ecosystems.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

10. CONCLUSION

The rising complexity of healthcare insurance claims documentation has posed a great problem to insurers to detect anomalies and fraudulent claims correctly. The common traditional methods of claims auditing are increasingly relying on manual reviews and rule-based processes, which are not able to process the increasing amount of unstructured documentation that is being produced within the current healthcare systems. The paper illuminates the promise of Natural Language Processing as a disruptive analytic technology that can enhance the understanding of the healthcare claims narrative and increase the ability to identify anomalies in insurance processes.

The suggested NLP-based framework provides a conceptual design of the integration of text analytics with anomaly detection models to improve claims auditing operations. The framework also allows providers to document and billing descriptions to be automated hence providing insurers with a better insight into loopholes that can suggest either errors, abuse or claims submissions in an effort to commit fraud. The framework also facilitates a more proactive claims review process through giving priority to high-risk claims to be reviewed further thus enhancing operational efficiency and risk management strategy.

In addition to its analytical works, the framework demonstrates the general tendencies of the digital revolution of healthcare insurance. By adding AI-based analytics to the claims processing processes, the regulatory compliance can be improved, and the financial integrity will be enhanced to facilitate a better management of insurance governance. With the further implementation of the digital technologies by the healthcare insurance organizations, the applications of the NLP-based claims analytics will become even more significant in the issues of the complex healthcare data environment.

Subsequent studies can also consider the use of superior machine learning models, language models, and real-world assertions sets to confirm and enhance the framework. This kind of research will help in further building of smart healthcare insurance systems that can enhance fraud detection, operational efficiency, and decision-making based on facts.

REFERENCES

1) Popowich, F. (2005). Using text mining and natural language processing for health care claims processing. *ACM SIGKDD Explorations Newsletter*, *7*(1), 59-66. https://doi.org/10.1145/1089815.1089824
2) Alkomah, B. (2022). *Abuse Detection in Medical Claims Using NLP and Deep Learning Techniques* (Master's thesis, University of Idaho). https://www.proquest.com/openview/9fb8a357d773cadf10fd916585a4c1c1/1?pq-origsite=gscholar&cbl=18750&diss=y
3) Nath, K. (2025, December). Natural Language Processing (NLP) for Improving Insurance Claims and Prior Authorization Processing. In *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 217-223). IEEE. https://doi.org/10.1109/ICAAIC64647.2025.11330756
4) Islayem, R., Gebreab, S., AlKhader, W., Musamih, A., Salah, K., Jayaraman, R., & Khan, M. K. (2025). Using large language models for enhanced fraud analysis and detection in blockchain based health insurance claims. *Scientific Reports*, *15*(1), 29763. https://www.nature.com/articles/s41598-025-15676-4
5) Hassan, M., & Alam, M. (2025). Automating Healthcare Claims Processing with Supervised and Unsupervised AI Models. *Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems*, *9*(2), 1-11. http://polarpublications.com/index.php/JABADP/article/view/2025-02-04
6) Taneja, A. (2023). Using NLP and AI to Automate Medical Coding and Insurance Claims on Cloud Systems. *International Journal of Emerging Research in Engineering and Technology*, *4*(4), 33-42. https://doi.org/10.63282/3050-922X.IJERET-V4I4P104
7) Rehman, R., Mazhar, A., Faheem, A., Matloob, I., Khan, S. A., Shaukat, S., ... & Siddiqui, S. (2025). AI Driven Framework for need based Insurance Plans Generation and Anomaly detection Using Deep Learning techniques. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3583562
8) Bauder, R. A., Khoshgoftaar, T. M., Richter, A., & Herland, M. (2016, November). Predicting medical provider specialties to detect anomalous insurance claims. In *2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI)* (pp. 784-790). IEEE. https://doi.org/10.1109/ICTAI.2016.0123

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

9) Saddi, V. R., Gnanapa, B., Boddu, S., & Logeshwaran, J. (2023, December). The Role of Natural Language Processing in Detecting Insurance Fraud. In *2023 4th International Conference on Communication, Computing and Industry 6.0 (C216)* (pp. 1-6). IEEE. https://doi.org/10.1109/C2I659362.2023.10430658
10) Matsuda, S., Ohtomo, T., Tomizawa, S., Miyano, Y., Mogi, M., Kuriki, H., ... & Watanabe, S. (2021). Incorporating unstructured patient narratives and health insurance claims data in pharmacovigilance: natural language processing analysis of patient-generated texts about systemic lupus erythematosus. *JMIR Public Health and Surveillance*, *7*(6), e29238. https://doi.org/10.2196/29238
11) Mazumder, S. A., & Rhaman, M. A. (2024). Patient care and financial integrity in healthcare billing through advanced fraud detection systems. *Academic Journal on Business Administration, Innovation & Sustainability*, *4*(2), 10-69593. https://dx.doi.org/10.69593/ajbais.v4i2.74
12) Umakanth, A. A. M. (2025). Generative AI in Healthcare Claims Processing and Fraud Detection: Transforming Insurance Workflows. *Journal of Computer Science and Technology Studies*, *7*(6), 739-745. https://doi.org/10.32996/jcsts.2025.7.87
13) Fursov, I., Zaytsev, A., Khasyanov, R., Spindler, M., & Burnaev, E. (2019). Sequence embeddings help to identify fraudulent cases in healthcare insurance. *arXiv preprint arXiv:1910.03072*. https://doi.org/10.48550/arXiv.1910.03072
14) Palavali, D. R., Pothireddy, S., Elumalai, D. K., & Nagaraja, M. (2026, February). AI-Driven Claims Adjudication: Optimizing Healthcare Systems with Machine Learning and Deep Neural Networks. In *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICAIC67076.2026.11395689
15) Owen, O. (2025). Deep Learning Approaches for Detecting Anomalous Billing Patterns in Healthcare Claims. *Available at SSRN 5418578*. https://dx.doi.org/10.2139/ssrn.5418578
16) Machireddy, J. R. (2023). Automation in healthcare claims processing: Enhancing efficiency and accuracy. *International Journal of Science and Research Archive*. https://doi.org/10.30574/ijsra.2023.9.1.0435
17) Mishra, A. Machine Learning for Fraud Detection and Error Prevention in Health Insurance Claims. *IJAIDR-Journal of Advances in Developmental Research*, *14*(1). https://doi.org/10.5281/zenodo.14615792
18) Li, W., Ye, P., Yu, K., Min, X., & Xie, W. (2023). An abnormal surgical record recognition model with keywords combination patterns based on TextRank for medical insurance fraud detection. *Multimedia Tools and Applications*, *82*(20), 30949-30963. https://link.springer.com/article/10.1007/s11042-023-14529-4
19) Suman, J. V., Mahammad, F. S., Sunil Kumar, M., Sai Chandana, B., & Majji, S. (2024). Leveraging natural language processing in conversational AI agents to improve healthcare security. *Conversational Artificial Intelligence*, 699-711. https://doi.org/10.1002/9781394200801.ch38
20) Bhatt, S., & Garg, G. (2025). NLP for Fraud Detection and Security in Financial Documents. In *Transformative Natural Language Processing: Bridging Ambiguity in Healthcare, Legal, and Financial Applications* (pp. 131-155). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-88988-2_6
21) Arora, S., Pandey, M., Arora, M., Gupta, K., Sharma, V., & Nagpal, L. (2024). Digitization of health insurance documents for the cashless claim settlement using intelligent document management system. *Procedia Computer Science*, *235*, 1319-1331. https://doi.org/10.1016/j.procs.2024.04.125