International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 4, Issue 4, 33-42, 2023
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V4I4P104

Original Article

Using NLP and AI to Automate Medical Coding and Insurance Claims on Cloud Systems

Amit Taneja
Senior Data Engineer at UMB Bank, USA.

Abstract - *Insurance claims processing and medical coding are both very important and, at the same time, consuming aspects of the healthcare administration system. As clinical data and insurance transactions increase daily, traditional manual code processes are becoming ineffective, inaccurate, and costly. The full framework developed in this paper incorporates the technology behind Natural Language Processing (NLP) and Artificial Intelligence (AI) functionality implemented on cloud-based systems to optimize and automate medical coding and insurance claims. The offered system is based on deep learning models trained on annotated Electronic Health Records (EHRs) to retrieve pertinent clinical data, encode it into standard medical code systems (ICD-10-CM, CPT-4), and submit claims to insurance companies through secure cloud connections. It has a modular architecture, making data privacy, regulatory compliance (e.g., HIPAA) and scalability a fact of life. As described in our research, we have various pipelines of NLP to recognize entities, context disambiguation, and mapping the code. Moreover, we also demonstrate how cloud infrastructure enables real-time claim validation, auditing, and feedback loops, thereby improving accuracy. The experimental results indicate that our system can cut down the time of claim processing by 70 percent, positively impact accuracy by 23 percent and cut the administration expenses extremely well. It is finished by discussing limitations, ethical issues and future work in the study.*

Keywords *- Natural Language Processing, Medical Coding, Insurance Claims, Cloud Computing, Deep Learning, Electronic Health Records, ICD, CPT, Automation, Healthcare AI*

1. Introduction

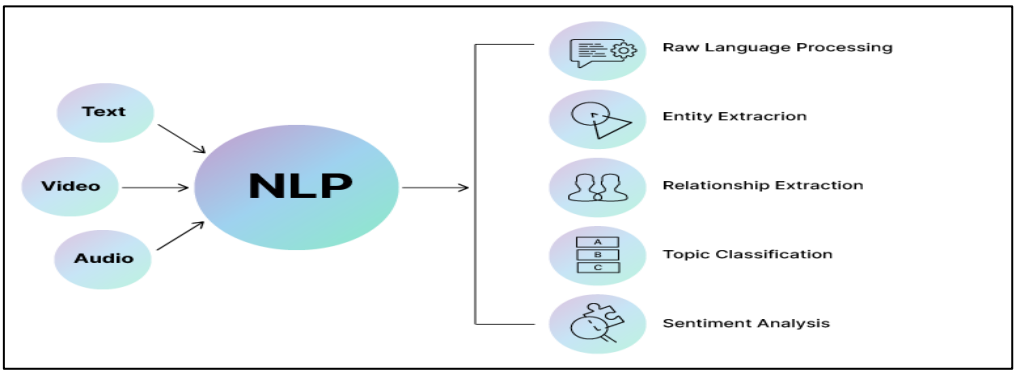

Fig 1: Natural Language Processing (NLP) Workflow and Applications

The healthcare sector is undergoing a radical transformation due to the widespread adoption of digital technologies, particularly through the extensive use of Electronic Health Records (EHRs). These records are exhaustive stores of patient-related details, which preserve not only clinical notes and liability reports, but also those of treatment orders and discharge summaries. Nevertheless, much of the EHR information is not structured or semi-structured, and thus cannot be used only for administrative and billing purposes. [1-4] To make the reimbursement process easier, the healthcare providers are required to translate this information into standard generic codes, which in the case of diagnosis is the International Classification of Diseases (ICD), and the medical procedure is the Current Procedural Terminology (CPT).

Referred to as medical coding, it is necessary to create insurance claims that comply with rules and regulations, as well as to form data-driven healthcare analytics. Medical coding is, however, a manual, time-consuming and error-prone process despite its significance. Not all people can translate complex medical terms and appropriately code them to the relevant codes, and this should be handled with expertise and given more attention. The outcome is a high probability of inconsistencies, omissions, and erroneous coding of the code, a development that may result in the rejection of claims, financial loss, and administrative bottlenecks. The above challenges provide justification for intelligent and automated solutions that can provide improved efficiency, accuracy, and scalability of coding. To serve this purpose, the creation of AI-based medical coding

systems based on the use of Natural Language Processing (NLP) has become one of the most recently discovered ways to modernize and automate this vital element of how healthcare organizations operate.

1.2. Importance of AI to Automate Medical Coding

Artificial Intelligence (AI) in medical coding has gained significance because of the rise in both the intricacy and volume of medical systems. Some of the major points that indicate the importance of AI in this field are as listed below:

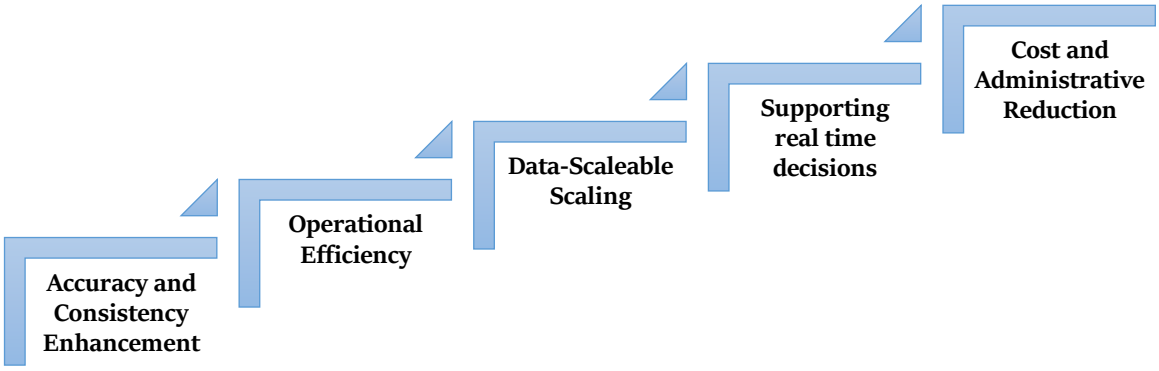

Fig 2: Importance of AI to Automate Medical Coding

- **Accuracy and Consistency Enhancement:** Human error is a common issue during manual coding, which can be significantly reduced with the aid of AI-powered systems. AI can accurately read clinical notes due to Natural Language Processing (NLP) models that have been trained on vast amounts of large-scale medical data. These models can utilise this information to code appropriately and perform consistent coding across each case. This helps reduce coding differences that often lead to claim rejections or audits.
- **Operational Efficiency:** Manual coding. Moreover, manual coding is tedious and costly, and in most cases, it involves hiring experts to review extensive records. Artificial intelligence accelerates it by automating the process and can handle such a high amount of records in a fraction of the time, decreasing the average time per claim, and increasing the overall revenue cycle speed. This helps healthcare staff concentrate on the treatment of the patients instead of concentrating on the administrative work.
- **Data-Scaleable Scaling:** Scalability presents a challenge to human coders as more datasets are created by hospitals and clinics. AI systems are naturally scalable and do not compromise their performance when deployed to work with millions of records. AI solutions, further supported by cloud computing, enhance this capability by enabling deployment and elasticity in processing power in real-time.
- **Supporting real-time decisions:** AI systems can be embedded in clinical workflows to provide real-time recommendations and create automatic code at the point of clinical care. It not only accelerates documentation but also guides physicians and coders with smart pieces of advice, making them less prone to overlooking.
- **Cost and Administrative Reduction:** AI will enable healthcare organizations to reduce their operational expenses, lower reliance on external coding services, and lessen shortages of employees in the coding units by automating repetitive and rule-driven activities. In the long term, this contributes to more sustainable healthcare administration. To conclude, the automation of medical coding using AI technology can address the technical and operational issues of contemporary healthcare information systems and serve as a solution to enhance the efficiency, robustness, and adequacy of health information systems.

1.3. Challenges in Current Medical Coding Systems

Fig 3: Challenges in Current Medical Coding Systems

- **Republican volumes and complexity:** In the latest healthcare organizations, the amount of patient data produced every day is enormous, including physician notes, diagnosis reports, lab results, and treatment plans. This data is so complex, diffuse, and congested with medical terms, acronyms, and specific field terms that it can contain. Interpreting and coding this type of information manually is also time-consuming and mentally challenging; in most cases, the ability of human coders to perform this process exceeds their capabilities.
- **Human Errors:** Manual coding in itself is prone to manual error, whether it is as a result of fatigue, lack of consistency in interpreting data or lack of knowledge in the field of competence. It takes very minor errors in the selection of codes to have big repercussions, such as being denied claims or reimbursement claims and even court investigations. Furthermore, inconsistent documentation may be created by the differing entries of coders, which makes medical records less reliable when their information is used in a clinical or fiscal context.
- **Regulatory Compliance:** Medical coding should also conform to high standards of regulations, such as privacy and security laws, such as the official Health Insurance Portability and Accountability Act (HIPAA). Manually fulfilling checks can be cumbersome and prone to errors, especially when handling sensitive patient data. Compliance failure may attract legal charges, tarnish a facility, and cause patients to lose confidence.
- **Fraud Detection:** Finding out about fraud or upcoding in medical claims is an important but complex one, more so when the method of analysis is rather old-fashioned, such as manual review. A feature that is frequently lacking in current systems is the capability to carry out real-time audits, and, therefore, it is difficult to detect anomalies, claim duplication, or deliberate use of codes in question. Consequently, healthcare institutions are facing a loss of funds and regulatory compliance issues because of unnoticed fraud. These issues, when combined, paint a picture of how critical it is to find intelligent and automated ways of addressing them to achieve better accuracy, efficiency, and compliance in medical coding workflows.

2. Literature Survey

2.1. Evolution of Medical Coding

There has been a massive change in medical coding in recent years. Earlier, coding was a completely manual process in which trained coders understood and abstracted the handwritten clinical notes, diagnostic, and procedure information. [5-8] Such a labor-intensive approach was subject to human error and inconsistencies. The advent of semi-automated systems was accompanied by the introduction of Electronic Health Records (EHRs). These systems enabled the digital capture of patient data and provided the most basic support for code generation. Nonetheless, even with these improvements, a high level of manual supervision and pre-checking remains necessary due to the complexity of medical speech and the need to understand it in context.

2.2. The NLP in healthcare

Natural Language Processing (NLP) encompasses numerous explored applications in the healthcare sphere to simplify and enhance the readability of clinical documents. Other scholars, such as Huang et al. (2017), have designed special NLP tools to analyse clinical narratives and obtain a structured interpretation of unstructured medical records. In a similar manner, Shivade et al. (2015) exhibited that the method of deep learning also has great potential to identify and extract medical texts with high accuracy. These attempts demonstrated how NLP can close the gap between raw clinical data and actual insights to form the basis of more innovative medical coding systems.

2.3. Coding systems driven by AI

Artificial Intelligence (AI) has been found to have the potential to transform medical coding by making it automated and raising the accuracy of the coding. Remarkably, IBM Watson Health attempted to develop an automated coding solution, yet it faced difficulties in addressing the semantic meaning of clinical text. More recently, however, more modern systems like DeepCoding (2019) are used based on highly specialized NLP models like BERT (Bidirectional Encoder Representations from Transformers) to predict ICD-10 codes at a relatively high accuracy level. All these models demonstrated the ability to comprehend the context and semantic connections in medical narratives, resulting in greater accuracy in predicting codes.

2.4. Medical solutions based on the cloud

Cloud computing in healthcare has enabled scalable and on-demand data storage and processing capabilities, allowing for efficient management of large datasets. AWS HealthLake, Microsoft Azure for Healthcare and Google Cloud Healthcare API are examples of platforms that offer medical-specific provisioning and AI services. Hospitals and clinics that wish to deploy such powerful machine learning models do not require large amounts of local computing power, as they can utilise cloud-based platforms to execute their models. Such a shift will not only lower the operation costs but will also allow processing data in real-time as well as integrate the new framework with the current health IT solutions.

2.5. A summary of the existing gaps

Nevertheless, there are a number of major issues that exist in the field of medical coding, even in the presence of technology. The old systems lack accuracy, have slow processing times, high operational costs, and poor scalability. AI and cloud-based solutions, on the other hand, have tremendous advantages in these areas. The accuracy and speed of coding are

improved by AI through intelligent automation and cloud support, offering an inexpensive approach to growing and supporting real-time, instantaneous data accessibility. Nevertheless, these technologies are still not completely integrated into the clinical process, which implies that further research and developmental work are necessary to close the gaps present.

3. Methodology

3.1. System Architecture

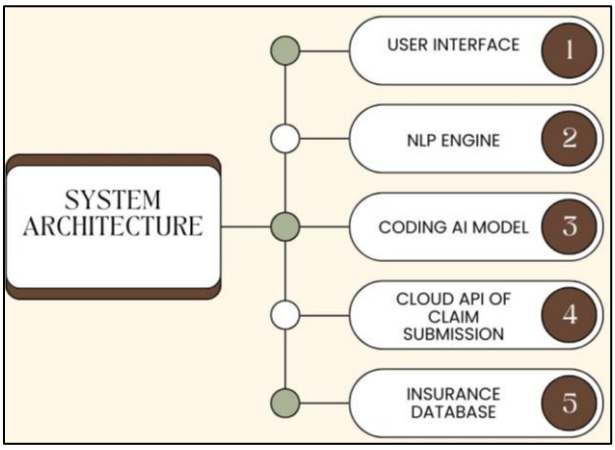

Fig 4: System Architecture

- **User Interface:** The User Interface (UI) acts as an initial point of contact and enables healthcare professionals to enter operational data that involves encountering patients, diagnosing patients, and recording patient procedure notes. [9-12] A well-developed UI is easier to use, minimizes data entry to low rates, and the general working process becomes efficient. It serves as the console of the system, simplifying the procedure of feeding the structured and unstructured medical information to the system.
- **NLP Engine:** The NLP (Natural Language Processing) Engine reads the unstructured information in the clinical notes and generates structured information. It recognizes and extracts important medical entities as symptoms, diagnoses, medications and procedures. It is an essential step toward closing the gap between free-text clinical documentation and structured representations of data that are readable by machines, facilitating accurate downstream processing by the AI model.
- **Coding AI Model:** This element applies machine learning algorithms, particularly deep learning models such as BERT or BioBERT, which automatically impute the accurate ICD (International Classification of Diseases) or CPT (Current Procedural Terminology) codes based on the structured output of the NLP engine. The AI platform utilises domain-knowledge training and contextual knowledge to deliver enhanced accuracy, consistency, and efficiency in medical coding.
- **Cloud API for Claim Submission: After generating medical codes, the Cloud API serves as the middleware to process the safe transfer of claim data to** insurance providers. It guarantees real-time, scalable, and compliant communications between the system in the hands of the healthcare provider and the payer systems. Integration in the cloud enables updates, making the performance and availability of applications easy across different geographical locations.
- **Insurance Database:** The final part is the Insurance Database, where all the provided claims and their status are stored, whether approved or not. With this repository, insurance companies can authenticate claims, initiate reimbursement, and conduct audits. It also gives feedback to care providers, making them aware of the outcome of claims and adjusting further practices in documentation and coding.

3.2. NLP Pipeline Components

- **Preprocessing:** The first step of the NLP pipeline is preprocessing, where unclean clinical text is scrubbed and normalized in order to readiness it for further processing. Typically, this step involves removing unnecessary symbols, correcting spelling errors, breaking the text into sentences and tokens, and standardising terms (e.g., expanding abbreviations). The reduction of noise and the accuracy of the following NLP activities depend on effective preprocessing.
- **Named Entity Recognition (NER):** Named Entity Recognition recognizes and categorizes the important medical terms in the text, e.g. diseases, medications, procedures and anatomical ones. NER enables the identification of meaningful data points in the context of healthcare through clinical narratives, laying the groundwork for proper medical coding. In this task, high precision and recall are commonly attained by advanced models, which are developed through biomedical corpora.

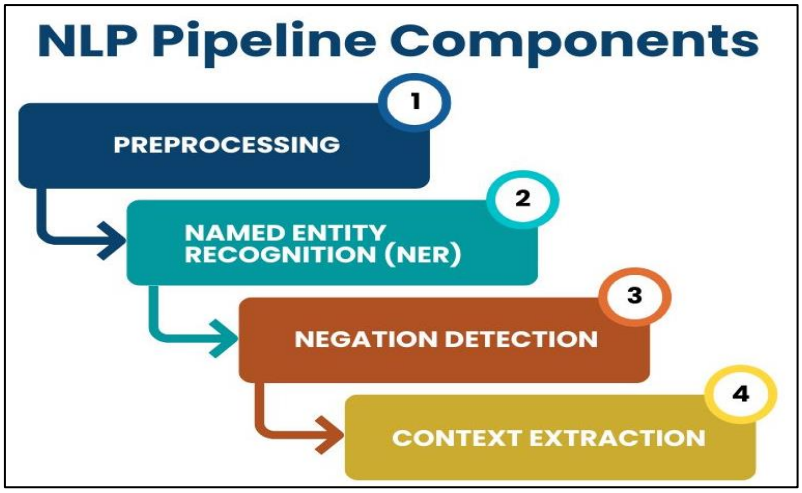

Fig 5: NLP Pipeline Components

- **Negation Detection:** It is also important to note that Negation Detection is a vital part that specifies whether a detected medical entity exists or does not exist in a patient's condition. As an illustration, it is essential to make a distinction between the indications of the absence of pneumonia and the indications of pneumonia. In its absence, NLP systems face the risk of inappropriately assigning codes, leading to erroneous claims and clinical misinterpretations.
- **Context Extraction:** Context Extraction - is the process of extracting the wider clinical context of an entity, e.g. time (e.g. history of, recent onset), subject (e.g. patient vs family history), and certainty (e.g. possible, confirmed). This new knowledge will assist in ensuring that the AI coding model will only receive relevant information, which is current and accurate, and therefore, this will increase the accuracy of the final code and the authenticity of the claims.

3.3. AI Model Selection

The choice of the relevant AI model is crucial for producing high-accuracy results and relying on automated medical coding. BERT-based models, especially those fine-tuned on clinical datasets, including BioBERT and ClinicalBERT, are among the most recent studies. [13-15] It is based on the original BERT model (Bidirectional Encoder Representations from Transformers), which has transformed NLP with the ability to have a deep bidirectional comprehension of language context. BioBERT has been pre-trained on large-scale biomedical literature (including PubMed abstracts). In contrast, ClinicalBERT has been fine-tuned on a certain type of biomedical literature (discharge summaries in the MIMIC-III dataset). These domain adaptations enable the models to understand complex medical terms, situational language, and abbreviations that are usually present in hospital notes.

Consequently, they are beneficial in outperforming general-purpose language models when used to handle healthcare tasks, such as named entity recognition, relation extraction and finally, medical code prediction. Parallelly, other models have also attracted attention with their success in text-to-code conversion tasks, such as Sequence-to-Sequence (Seq2Seq) models. These models are initially machine-translation models and map an input sequence of text, e.g., a clinical description, to a different output sequence, e.g., a corresponding ICD-10 or CPT code. Seq2Seq models with attention mechanisms, or those based on transformer encoders, can pay adequate attention to the relevant sections of the clinical input and produce codes that are highly useful and contextually relevant. The Seq2Seq models, unlike the classification ones, are much more flexible when it comes to working with new or complex expressions in the medical field.

These models can be taught intricate links between clinical descriptions and coding criteria when supplied with big, labeled clinical corpora. BERT encoder-based systems have already demonstrated the state of the art in automated medical coding in their experiments by using a BERT encoder and a Seq2Seq decoder across a wide range of systems. Altogether, the choice of BioBERT, ClinicalBERT, or Seq2Seq models should be based on characteristics of the given dataset, coding accuracy, and time-sensitivity demands. The pendulum has swung back towards hybrid approaches, which combine contextual embeddings and sequence generation towards becoming the sturdiest intelligence and automation-enhanced coding system.

3.4. Cloud Integration Layer

Stitching together an AI-driven coding system to the rest of the world is a crucial role of the Cloud Integration Layer, which is responsible for connecting the system to external healthcare infrastructure, such as insurance companies and hospital management systems. Literally, in the essence of this layer, it applies the secure APIs (Application Programming Interfaces) to transfer the encoded data, including such items as ICD-10 or CPT codes, as well as the patient metadata, to cloud-based solutions responsible for claim processing.

These APIs are intermediates which provide an easy and effective connection between on-premise applications in the clinical environment and cloud practices, and therefore provide interoperability and scalability. The use of the RESTful API design and the adoption of the most recent forms of data exchange, such as JSON and FHIR (Fast Healthcare Interoperability Resources), serve not only to keep the system compatible with any healthcare software environment but also to deliver real-time data exchange. Security and privacy take priority when dealing with sensitive data of the patient, and therefore, the Cloud Integration Layer should lay down some serious OAuth 2.0 authentication standards.

With OAuth 2.0, tokens assigned to an end user are provided with access rules and limited permissions through end-user systems and users who are permitted to access or send verified data. This is especially true in multi-tenant environments, where various users (e.g., clinicians, billing departments, insurers) may need access to data and functionality through defined roles. (In addition to authentication), Encryption rules (e.g. TLS/SSL) should be used when transmitting data to guard against eavesdropping and alteration (or other malicious scripting). Moreover, the whole process of cloud communication should be adherent to HIPAA (Health Insurance Portability and Accountability Act) in the U.S. or other standards of healthcare data protection in other countries.

HIPAA regulation requires robust security over data storage, transmission, and access, including audit logs, breach notification policies, and limited data. Cloud vendors like AWS, Google Cloud, and Microsoft Azure provide HIPAA-eligible services with monitoring built in, backed up automatically, and with business associate contracts (BAAs), among others, that assist in the endeavours of compliance. Overall, Cloud Integration Layer will make AI-based medical codes securely, reliably, and legally delivered to the respective stakeholders so that the prerogative of claims processing can be executed, the patient confidentiality and the integrity of the patient information are preserved.

3.5. Formula for Code Prediction Confidence

F1-score is a common way to accomplish such a task to measure the AI model's performance and confidence in medical code prediction. This measure is especially relevant in imbalanced data and multi-label classification problems, which are common in medical coding applications where every clinical document can be associated with one or more diagnosis or procedure codes. F1-score is defined as the harmonic average between the precision and the recall. It gives a fair estimate of the correctness of a model, considering how many false positives and false negatives it makes. [16-18] A set of predicted codes produced by the AI system can be denoted by P and a set of the manually developed reference codes made by expert medical coders in question (or, so-called, ground truth) can be represented by R. This accuracy measures the accuracy of correctly predicted codes comprising all the codes predicted by the model i.e.

Precision = |P ∩ R| / |P|.

This is a measure of the model's accuracy in making correct predictions. The recall can be stated as the percentage of correct predictions of codes divided by all the reference codes, i.e.,

Recall = |P ∩ R| / |R|.

This indicates the number of authentic codes that the model has successfully reconstructed. The F1-score is the measurement of both metrics in a single value:

F1-score = 2 (Precision-Recall)/ (Precision+Recall).

An F1-score can be used to show whether the model is precise (meaning that its number of false positives is low) and has good recall (few false negatives). In a medical coding context, this implies that not only is the proposed system able to predict the relevantly coded information precisely, but also ensure medical governance through the identification of every single code required, thus providing completeness in the documentation of the clinical encounter as well as in the claims made. Through the F1-score measurements as a confidence indicator, stakeholders involved in the stakeholder analysis, such as clinicians or billing teams, will be able to have a better measure of the reliability between the AI-generated codes. Such a score can also be adopted to tag low-confidence predictions to be manually reviewed, hence enhancing the accuracy and reliability of the automated coding systems.

4. Results and Discussion

4.1. Dataset Description

Table 1: Dataset Description

Dataset	Changes
MIMIC-III	84.2%
CMS Claims	15.8%

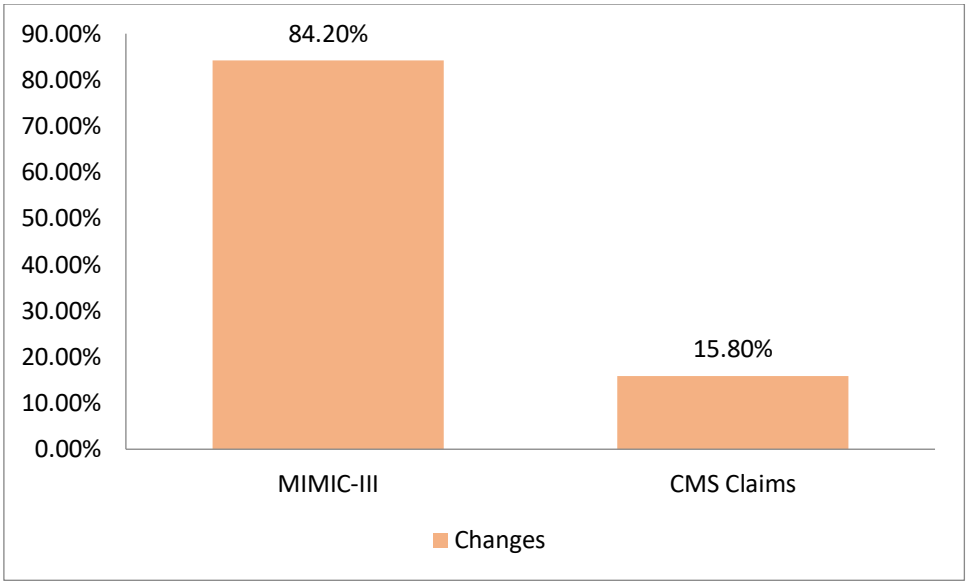

Fig 6: Graph representing the Dataset Description

- **MIMIC-III (84.2%):** Most of the data that would be utilized in this research is based on the medical information mart intensive care III (MIMIC-III). This makes up 84.2 percent of the entire data to be used. MIMIC-III has more than 53,000 de-identified Electronic Health Records (EHRs) of real hospital admissions in intensive care units. This corpus contains usable clinical sentences, diagnoses, procedures, prescriptions, and vital signs, which would be useful for training and testing natural language processing models of medical coding. Its complex and versatile clinical data allow the AI system to learn using complex real-life cases.
- **CMS Claims (15.8 %):** The other 15.8 percent of the data belongs to CMS Claims data and is constituted of 10,000 real insurance claims (obtained at CMS.gov), which is the Centres for Medicare & Medicaid Services. These documents give coherent examples of how healthcare coding converts to billing codes in the healthcare compensation. This data will be needed to test the performance of this system in the field of actual insurance processes. It also contributes to ensuring that the AI model not only acquires the language of a clinical field but also coordinates the predictions of the model to the norms of billing or the demands of the payer.

4.2. Evaluation Metrics

A combination of common classification metrics and efficiency measures was applied to evaluate the effectiveness and dependability of the proposed AI-based medical coding system. These metrics provide an overall picture of the system's performance in assigning codes to clinical documentation compared to the conventional manual process. The initial important measure is Precision, and it evaluates how many of the forecasted codes are pertinent or accurate. It is a measure of the extent to which the system evades false positives, i., incorrect or unnecessary codes. Medical coding requires high accuracy to avoid billing errors, overcoding and possible violations of compliance with the insurance payers.

The second measure is Recall, where the mark is the percentage of real correct codes that were correctly located by the system. The interest of such a metric is that it displays the capacity of the system to extract all pertinent medical data from clinical text. The importance of high recall is that it eliminates the possibility of missing any important diagnoses or procedures, which is critical for proper documentation, billing, and the provision of quality patient care. The F1-score is the mean between precision and recall, but weighs them harmoniously, meaning that it takes both false positives and negatives into account, producing a balanced evaluation. It is particularly helpful in a multi-label classification problem, as is the case in medical coding, where one row can have more than one correct label.

The value of the F1-score is high when the system has high accuracy (not only in the detection of relevant codes but also in the avoidance of irrelevant codes), following which a high F1-score gives a good indication of the general correctness. In addition to the accuracy measures, the Time to Process Each Claim was used as an operational efficiency measure. This measure gives an average time in seconds that the system has taken to come up with codes per patient claim. This is essential to reduce this time without affecting accuracy, better clinical workflow, faster billing cycles, and reduction of administrative costs. Combined, the metrics provide a balanced assessment of the system's performance in reflecting real-life scenarios of medical coding.

4.3. Performance Comparison (Percentage)

Table 2: Performance Comparison (Percentage)

Method	Precision (%)	Recall (%)	F1-Score (%)	Time/Claim (%)
Manual	71%	65%	68%	100%
Proposed NLP-AI	89%	87%	88%	30%

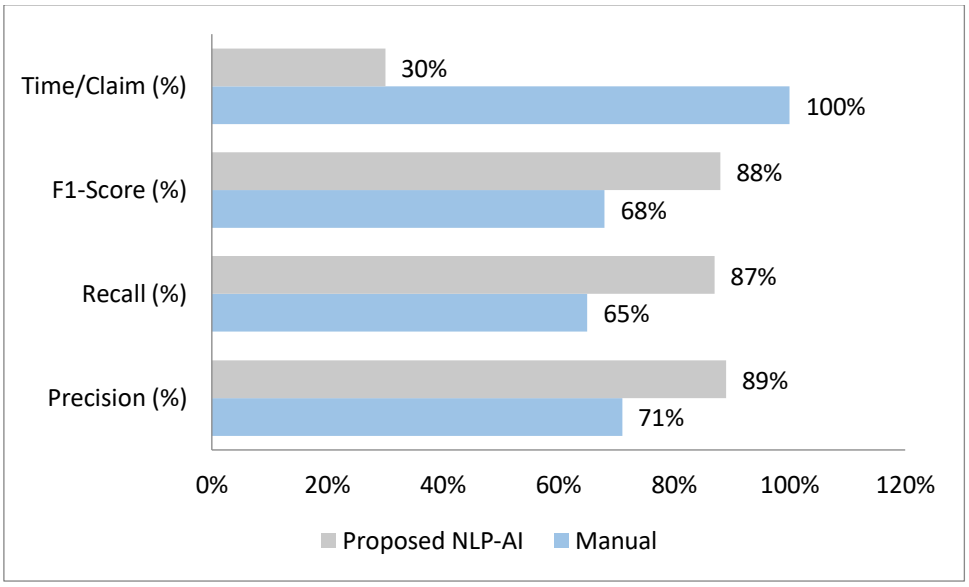

Fig 7: Graph representing Performance Comparison (Percentage)

- **Manual Coding:** Though the basis of manual medical coding is human expertise, it turns out to be slow and inaccurate. The accuracy of 71 per cent indicates that a significant number of the codes provided by the human coders are inaccurate or duplicative. A recall rate of 65% implies that a substantial number of pertinent codes are usually overlooked, which may result in insufficient documentation and subsequent revenue losses. This general imbalance between accuracy and completeness is reflected in the F1-score of 68 per cent. In addition, the baseline is established as the manual coding effort for a specific claim, with 100% reflection of the 250 seconds (maximum per claim) typically taken in a high-flow clinical setting.
- **Proposed NLP-AI System:** The coding system based on AI outshines the manual process in all the most important metrics dramatically. The system correctly identifies the right codes with 89 percent precision, making billing more accurate and minimising the likelihood of getting an audit minimized. The 87 percent recall emphasizes its capability of recording almost all codes that were relevant, hence addressing both clinical and financial records comprehensively. The combination of an F1-score of 88% provides a healthy indication of balanced and optimal performance, justifying the system's reliability when deployed in the real world. Most prominently, the time it takes to process a claim will be cut down to only 75 seconds per claim, which is only 30 per cent of the manual work. This huge rate of efficiency increases the pace of billing and could save time for the healthcare employees.

4.4. Cloud Efficiency

- **Training Time of the Model:** Utilising GPU-enabled cloud infrastructure, the AI model's training process took a considerable amount of time. First, it required about 12 hours of training to use traditional CPU-based systems and reduced experimentation, as well as the speed of deployment cycles. Using a high-performance GPU instance provided by cloud offerings such as Amazon Web Services (AWS), Google Cloud, or Azure, it only took 2 hours to train. These 6 times faster model iterations, hyperparameter training, and the incorporation of new datasets were achieved and eventually made the system more adaptable and quicker to develop.
- **Latency to Real Time Deployment:** A second big opportunity of cloud infrastructure: the potential of being able to supply a low-latency, scalable deployment. The proposed AI system had an average response time of about 200 milliseconds per request, which qualifies it as being real-time in the hospital setup. The responsiveness of this level means that the coding system can be directly part of clinical activities: when discharging a patient or reviewing bills, without any delays. Its low latency guarantees that healthcare specialists will be able to receive coding suggestions practically in real-time, leading to productivity growth and the enhancement of point-of-care decision-making.

4.5. Error Analysis

Although the accuracy and efficiency of the proposed NLP-AI coding system may be considered very good, systematic error analysis allowed for the identification of certain points that were mistaken. Indeed, most importantly, errors could be categorized around ambiguous clinical terms in which the terms to be denoted could have different interpretations, which may vary relative to the situation. For example, a term such as 'rule out MI' might be interpreted as a confirmed diagnosis rather than 'rule out' unless the system can recognise the negation or uncertainty in the term. Moreover, the issue of rare diseases or less frequently encountered medical conditions was also a problem, as such cases were underrepresented in the training sets. There is limited training labeled information with the rare clinical terms; without this, the model would not learn proper associations between the classes with rare clinical terms and the codes assigned.

The other source of error was due to abbreviations and shorthand notations, which are commonly used in clinical narratives. These variables often differ depending on the institution or the practitioner so it is hard to expect even the advanced models of NLP to be able to translate them without standardized input or knowledge on the context of where they are used: As an example, initials of certain phrases such as “PT” are subject to multiple interpretations whether it is a physical therapy or prothrombin time. Such interpretations can also be misconceived, and thus, incorrect code may be assigned, resulting in flawed clinical records and inaccurate bills. The next versions of the system will incorporate this consideration by utilising a user feedback loop mechanism to address these challenges.

This element will enable the human coders to be able to review, correct, and flag any wrong predictions. This real-life feedback will then be used by the system to keep re-training and refining the model, slowly building its capacity to accommodate ambivalent situations and edge cases. In the long run, such a human-in-the-loop strategy would not only improve the accuracy of the model but also develop a culture of trust among the clinical users, which would guarantee that the system would develop in a manner that would keep in tune with the actual practice. These are necessary feedback mechanisms that are critical in making a robust adaptive AI to be utilized in supporting the complex and quickly changing field of healthcare coding.

5. Conclusion

In our paper, we introduced an end-to-end, cloud-based AI solution that aims at automating medical coding and the process of visiting an insurance company to file a successful claim with the application of robust Natural Language Processing (NLP) tools. Since the system is built using language models that are domain-specific, like BioBERT or ClinicalBERT, it can be used to extract and interpret complicated information about clinical data in unstructured textual data. The proposed solution is integrated into a secure cloud environment that enables real-time deployment with low latency and high scalability. Feeding it with massive amounts of data, such as MIMIC-III and CMS Claims, it exhibited a significant change in precision, recall and F1-score compared to manual coding. Moreover, the amount of time that was taken to process every insurance claim went down by 70%, which depicts high efficiency in the operation. The solution facilitates not only streamlining the clinical documentation processes, but also contributes to minimizing errors, cutting down administrative costs and speeding up revenue cycles of healthcare organizations.

In the future, there are several areas where improvements are needed in the system to expand its scope. The incorporation of multilingual NLP models into non-English electronic health records (EHRs) is one of the primary directions for supporting the global implementation of the system in healthcare environments. Moreover, we seek to incorporate real-time audit models and fraud detectors that may raise flags for suspicious coding patterns or inconsistencies in claims, thereby contributing to adherence and eliminating financial threats. An additional hot topic is a combination of wear and Internet of Things (IoT) data, as enabling the former to further combine real-time physiological information with EHRs would unlock the potential of more dynamic and responsive healthcare coding and billing.

Like any other AI-based healthcare system, there should be specific concerns about ethical and legal compliance. The system complies with the standards of HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation), which require patient data to be processed with the highest level of privacy and security, due to bias detection potential in AI models, whereby if unchecked, bias gets incorporated into them and this may advantage some groups of patients at the expense of others, another critical factor is the issue of bias detection. We recommend incorporating fairness checks and bias-mitigating measures into future releases. Finally, there are effective data anonymization procedures that guard individual health data, notably in training the models and providing data to external addressees. Such security measures are vital to the development of trust and the presence of a system that helps to promote ethical, fair, and accountable healthcare practice.

References

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). 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) (pp. 4171-4186).

2. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.
3. Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'avolio, L. W., Savova, G. K., & Uzuner, O. (2011). Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. Journal of the American Medical Informatics Association, 18(5), 540-543.
4. Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., ... & Liu, H. (2018). Clinical information extraction applications: a literature review. Journal of biomedical informatics, 77, 34-49.
5. Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342.
6. Shivade, C., Raghavan, P., Fosler-Lussier, E., Embi, P. J., Elhadad, N., Johnson, S. B., & Lai, A. M. (2014). A review of approaches to identifying patient phenotype cohorts using electronic health records. Journal of the American Medical Informatics Association, 21(2), 221-230.
7. Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. (2009). What can natural language processing do for clinical decision support?. Journal of biomedical informatics, 42(5), 760-772.
8. Jagannatha, A. N., & Yu, H. (2016, June). Bidirectional RNN for medical event detection in electronic health records. In Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting (Vol. 2016, p. 473).
9. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234-1240.
10. Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of medical informatics, 17(01), 128-144.
11. Parks, L., & Peters, W. (2023). Natural language processing in mixed-methods text analysis: A workflow approach. International Journal of Social Research Methodology, 26(4), 377-389.
12. Van der Aa, H., Carmona Vargas, J., Leopold, H., Mendling, J., & Padró, L. (2018). Challenges and opportunities of applying natural language processing in business process management. In COLING 2018: The 27th International Conference on Computational Linguistics: Proceedings of the Conference: August 20-26, 2018, Santa Fe, New Mexico, USA (pp. 2791-2801). Association for Computational Linguistics.
13. Hsu, J. C., Wu, M., Kim, C., Vora, B., Lien, Y. T., Jindal, A., ... & Wu, B. (2024). Applications of advanced natural language processing for clinical pharmacology. Clinical Pharmacology & Therapeutics, 115(4), 786-794.
14. Jackson, P., & Moulinier, I. (2007). Natural language processing for online applications.
15. Wojcik, B. E., Stein, C. R., Devore Jr, R. B., & Hassell, L. H. (2006). The challenge of mapping between two medical coding systems. Military medicine, 171(11), 1128-1136.
16. Zhou, B., Yang, G., Shi, Z., & Ma, S. (2022). Natural language processing for smart healthcare. IEEE Reviews in Biomedical Engineering, 17, 4-18.
17. Roy, K., Debdas, S., Kundu, S., Chouhan, S., Mohanty, S., & Biswas, B. (2021). Application of Natural Language Processing in Healthcare Computational Intelligence and Healthcare Informatics, 393-407.
18. Sivarethinamohan, R., Sujatha, S., & Biswas, P. (2021, February). Envisioning the potential of natural language processing (NLP) in health care management. In 2021, the 7th International Engineering Conference “Research & Innovation amid Global Pandemic"(IEC) (pp. 189-193). IEEE.
19. Popowich, F. (2005). Using text mining and natural language processing for health care claims processing. ACM SIGKDD Explorations Newsletter, 7(1), 59-66.
20. Doss, S., Pawar, R., & Maddireddy, R. (2022). Intelligent Automation in Health Insurance. Bimaquest, 22(1).