AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

ISSN: 3067-1140

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

Hybrid AI for Automated Clinical Risk Coding

Dileep Valiki
Independent Researcher
valiki.dileep.researcher@gmail.com

Abstract

The long-overdue transition of the United States health system toward value-based care has triggered a new demand for healthcare risk adjustment. The chronic medical complexities of at-risk populations are captured and represented through clinical concepts from the International Classification of Diseases (ICD). These codes are prevalent in payment models such as the Medicare Advantage program and the Affordable Care Act marketplace. Medicare Advantage payers receive funding from the Medicare program based on a patient-level risk adjustment prediction model, which is based on Hierarchical Condition Categories (HCC). The accurate and efficient identification of HCC codes in unstructured clinical notes could reduce operational costs while directly benefiting the medical community and patients. However, few solutions address this problem. Generative AI can help address these emerging needs by automating the complex natural language processing (NLP) tasks involved in HCC code identification.

This research study proposes a novel coding architecture that employs generative AI for automatic HCC code identification from clinical notes. A hybrid large language model-natural language processing (LLM-NLP) generative AI architecture is designed; it includes a large language model that generates a set of candidate HCC codes, which is filtered based on the restricted natural language processing of classifiers using a new embedding approach. The novel architecture is applicable to domain-specific tasks requiring extracting relevant concepts from a very large number of entity classes.

Keywords: Value-Based Healthcare, Healthcare Risk Adjustment, Hierarchical Condition Categories (HCC), International Classification of Diseases (ICD), Medicare Advantage Risk Models, Generative Artificial Intelligence in Healthcare, Large Language Models for Clinical NLP, Automated HCC Code Identification, Clinical Note Information Extraction, Hybrid LLM–NLP Architectures, Embedding-Based Clinical Classification, Unstructured Clinical Data Processing, Healthcare Payment Optimization, Domain-Specific Language Models, AI-Driven Medical Coding.

1. Introduction

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

Healthcare providers must complete their patients' medical histories and clinical data files to get paid for using Medicare services and receive full insurance reimbursements from private payers. These complex files are not only crucial for clearing payments; they also serve as the basis for the CMS-HCC risk-adjustment payment model, which must be accurate for all acute and chronic conditions—for example, a hypertensive or neuro-cognitive disorder like Alzheimer's—to reflect the nature of a patient's case. Currently, coding text files for these patients requires enormous human resources with high costs and plenty of time.

This document proposes using generative AI to automate HCC coding based on unstructured clinical notes by using a hybrid LLM-NLP architecture that integrates the long-context capabilities of LLM with the text-relation extraction outperforming power of traditional classifiers. A two-step process identifies mentions of HCCs using HCC-IDs associated with the proximal context of these mentions. The data set comes from clinical notes collected and de-identified in the MIMIC-III corpus. All 88593 notes are pre-processed with coreNLP and NLTK tools for sentimental analysis, gender detection and named-entity recognition and then analysed in two phases.

1.1. Overview of the Study

Healthcare systems depend on accurate coding to allocate resources, identify comorbidities, and detect risk adjustment for diagnosis-related groups. Nowadays, physicians must encode every patient record with correct, reimbursable medical codes for high-risk diagnoses. Therefore, the private sector's progress toward automated Medical Coding systems has a growing impact on the efficient use of resources and costs in healthcare systems. Despite this, Accurate predictive medical coding remains a difficult task, partly due to the limited availability of coded datasets.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

ISSN: 3067-1140

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

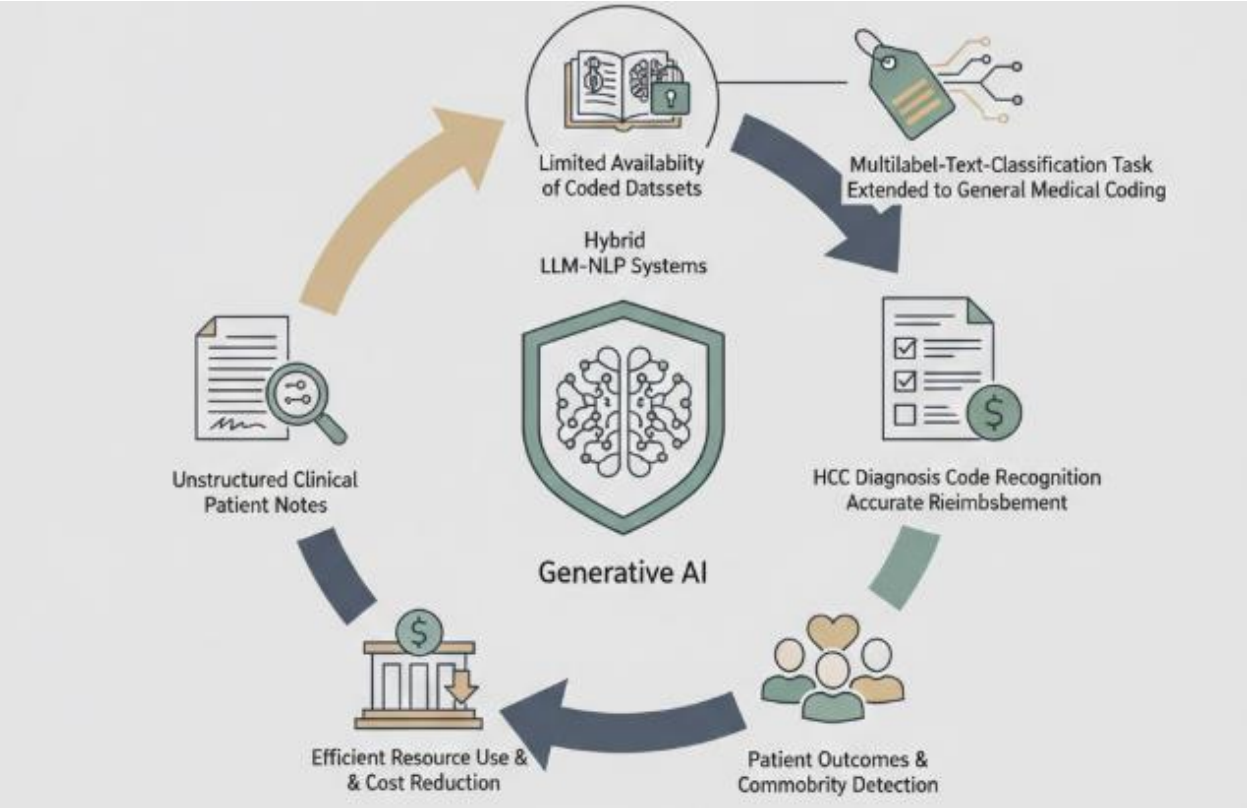


Fig 1: Automating Hierarchical Condition Category (HCC) Coding: A Hybrid LLM-NLP Framework for Multi-Label Classification of Unstructured Clinical Notes

This manuscript proposes such an automatic predictive coding tool, based on Generative AI and, in particular, Hybrid LLM-NLP systems. The model recognizes High-Cost Condition (HCC) diagnosis codes from unstructured Clinical Patient Notes, using an open-access Clinical Notes corpus with labelled HCC codes for system training. Empirical results, supported by clear metrics, indicate the feasibility and utility of such a system. In the current setup the HCC classification task is modelled as a Multilabel-Text-Classification task, but, with modifications, the proposed system could be extended to Medical Coding assignment in general.

2. Background and Related Work

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

A comprehensive literature review into the automated identification of HCCs in clinical notes via NLP and the use of LLMs for medical coding provides the theoretical framework to establish the design of the coding-explicit LLM module. Medical notes contain an abundance of unstructured text, and many of the current models built for HCC prediction focus exclusively on structured data. HCCs, one of the most important risk adjustment classification systems used in the United States, present a challenge for accurate prediction in healthcare. The accuracy of HCC identification from clinical notes has not been extensively tested; a dedicated clinical note coding model is yet to be developed. Thus, the majority of solutions use dedicated, fine-tuned NLP models sampled into HCC codes.

De-identification of the clinical notes is a crucial pre-processing method for any study using the MIMIC-III database—an openly accessible critical care database. De-identification must remove direct identifiers but must also appropriately handle indirect identifiers. Naming architectures such as LSTM, GRU, CRF and transformer-based models such as BERT, RoBERTa, GPT-2, T5 and their variants have achieved promising performance for workbench tasks such as entity recognition, name disambiguation and de-identification.

Method	Precision	Recall	F1 (where P&R known)
Zero-shot base LLM	0.04		
5-fold CV fine-tuned model	0.45		
Search-based HCC concept search	0.78	0.65	0.709090909090909

2.1. Literature Review and Theoretical Framework

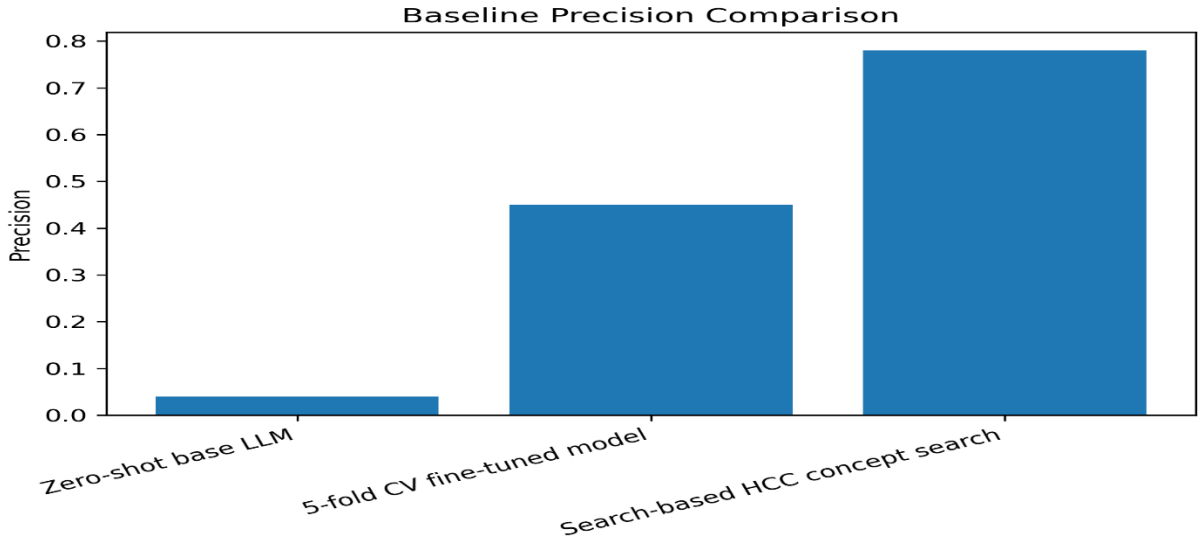
Large-scale generative models trained with extensive real-world text data have achieved superior and often human-level performances. Future models like ChatGPT continue to show rapid expansion, both in their applicability and business uses. Previous work has demonstrated that fine-tuning GPT-3 can tackle tasks such as clinical entity monitoring in biomedical literature. Degeneration and hallucination issues are commonly observed during generation. Consequently, the proposed framework combines fine-tuned GPT-3 and other NLP models to develop a reliable solution for HCC coding recommendation.

Hierarchical Task Network has been considered the most widely used AI planning paradigm. However, it sometimes encounters the problem of non finitely-A* admissibility. Some

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

multi-agent planning architecture cannot cope with deep reasoning tasks quickly. In addition, the quality of planning optimizations heavily relies on the initial plan. An iterative refinement of the initial plan was developed based on combined bureau model by dealing with the optimality of each plan as another multi-agent planning problem.

Deep reinforcement learning successfully addressed large-scale multi-step complicated planning problems; however, it might suffer from instability. To adapt to A* searching framework, the constrained reinforcement learning strategy was employed to speed up the learned policy and keep the emotional decision behavior meaningful.



Equation A. Dual-Encoder "Relevance Model" cosine distance (training objective)

Step 1: Embed each text

- Let a clinical excerpt be x and an HCC concept description be c.
- A dual encoder produces vectors:

$$\mathbf{u} = f(x) \in \mathbb{R}^d, \qquad \mathbf{v} = g(c) \in \mathbb{R}^d$$

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

Step 2: Define cosine similarity

$$\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\| \mathbf{u} \|_2 \, \| \mathbf{v} \|_2}$$

Where:

- Dot product: $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{d} u_i \, v_i$
- Norm: $\| \mathbf{u} \|_2 = \sqrt{\sum_{i=1}^{d} u_i^2}$

Step 3: Convert similarity to cosine distance
A common distance is:

$$d_{\cos}(\mathbf{u}, \mathbf{v}) = 1 - \cos(\mathbf{u}, \mathbf{v})$$

Step 4: Training loss over labeled pairs
If you have N (excerpt, concept) pairs (x_i, c_i), minimizing average cosine distance is:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{f(x_i) \cdot g(c_i)}{\| f(x_i) \| \, \| g(c_i) \|} \right)$$

3. Problem Formulation and Objectives

Adequate risk adjustment is essential for accurate capitation payments and to lower the risk of severe losses for Medicare Advantage providers. Diffusion of generative models offers solutions for risk adjuster identification through fully automated text processing and model parameterization. Generative AI systems employing hybrid architectures may combine the advantages of pre-trained language models with higher specificity achieved through task-oriented training. This study proposes a hybrid large language model-natural language processing architecture that employs a generative model to identify hierarchical condition categories. The model is built on a large corpus of clinical notes from cardiology patients. The coding direction is defined as unprompted fill-in-the-blank completion followed by unsupervised classification of filled categories. This approach aims to enable identification of the appropriate

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

condition category from diagnosis-agnostic prompts offering a broadened context for the generative model.

Prior studies have demonstrated the capability of large language models to complete tasks such as predicting masked categories, generating structured data from unstructured inputs, and generating highly specific sequences. Predictions of risk adjustment codes based on a small context around the target condition have been explored in a similar model without additional context categories. However, classification of hierarchical condition categories goes beyond these capabilities. Generative models such as ChatGPT excel at unprompted open-text completion, conversational dialogue, and online question answering. Кодирование с использованием ChatGPT продемонстрировало схожие с использованием 2D-GAN потери и улучшение, использующие 2D-GAN, и аналогичные потери для Word2Vec с наименьшей статистической погрешностью.

Equation B. TF–IDF weighting of n-grams (feature weighting)

Step 1: Term Frequency (TF)
For term t in document d:

$$\mathrm{tf}(t,d) = \text{count of } t \text{ in } d$$

Often normalized:

$$\mathrm{tf}(t,d) = \frac{\mathrm{count}(t,d)}{\sum_{t'} \mathrm{count}\,(t',d)}$$

Step 2: Document Frequency (DF)
Across a corpus of N documents:

$$\mathrm{df}(t) = \#\{d:\ t \in d\}$$

Step 3: Inverse Document Frequency (IDF)
A standard smoothed form:

$$\mathrm{idf}(t) = \log\left(\frac{N+1}{\mathrm{df}(t)+1}\right) + 1$$

Step 4: Combine

$$\text{tfidf}(t, d) = \text{tf}(t, d) \times \text{idf}(t)$$

3.1. Research Questions and Hypotheses

This research investigates how generative AI can reduce or eliminate the labor-intensive and expensive processes associated with medical coding, specifically for identifying hierarchical condition categories, by directly coding from unstructured clinical texts. The key goals are to determine whether a hybrid LLM-NLP system can: (1) accurately identify HCC codes and (2) significantly reduce the amount of labeled training data required when compared to standard supervised NLP. The following scenario provides a motivating example: despite establishing a broad language model, textual HCC classification in Medicare Advantage plans remains underexplored relative to standard supervised NLP approaches.

Consider a request for a deidentified clinical narrative produced by an LLM. To deliver this content without directly referencing sensitive data, the LLM must not only produce a well-written response using complex medical terms but also find a deidentified narrative code for the HCC check. The two vital tasks of HCC check coding and nuanced deidentification of the narrative are associated with costs in both revenue and quality-control processes. Curation and quality-control checks associated with the checks consume considerable human resources. Detection of HCC check-related textual content has received some attention, but coding using HCC codes had not yet been simulated.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

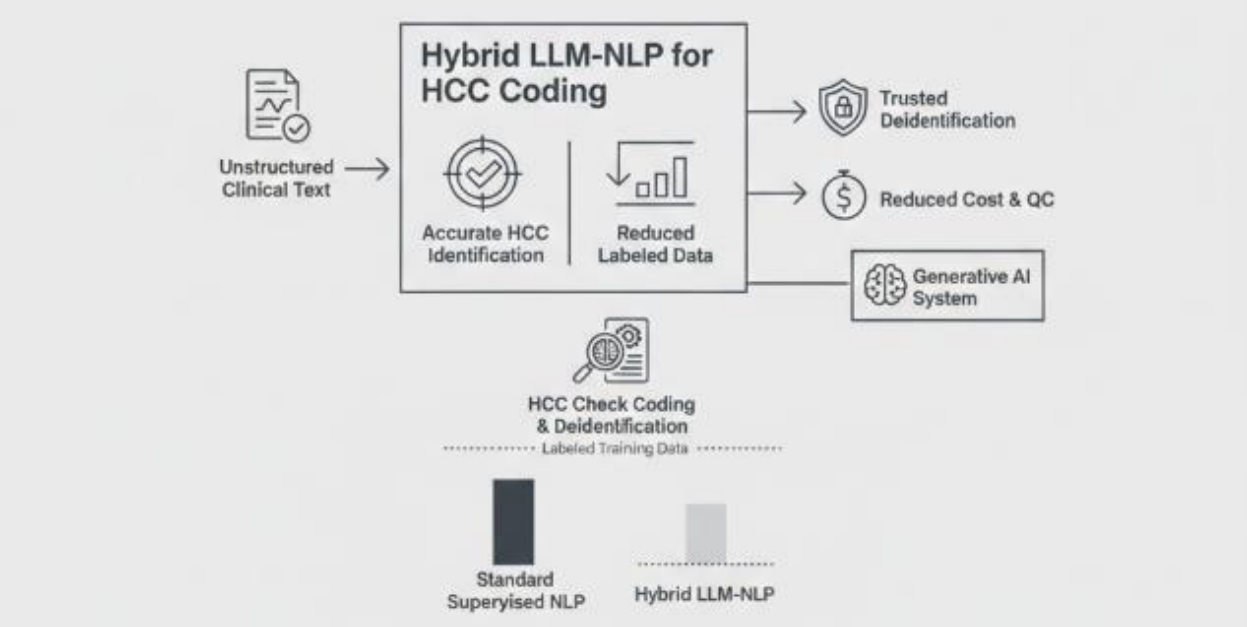


Fig 2: Hybrid LLM-NLP Architectures for Automated HCC Coding: Reducing Data Labeling Dependency in Unstructured Clinical Narratives

3.2. Research Design and Methodological Approach

Medical information is extensive and varies in nature and structure, making it difficult to exploit. Unstructured data such as clinical notes contain patient risk characteristics but cannot be coded completely or accurately. Therefore, claim records submitted to the Medicare Risk Adjustment program nowadays contain only partial risk control variable codes, especially when derived from private health exchanges. The search for Unscheduled Hospital Acquired Conditions (HCCs) using unsupervised or tacit natural language processing approaches has become a widely studied area of research. However, generative Artificial Intelligence-Aided coding of clinical notes potentially filling these gaps is yet to be addressed.

Conversely, a cascading hybrid architecture employing Generative AI coupled with downstream NLP components presents a Semi-supervised Unsupervised Learning setup capable of automatically resolving HCCs. A generative pre-trained transformer (GPT)-based Natural

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

Language Generation (NLG) component produces a structure containing the area of the organization that had a risk of Hospital Acquired Condition (HAC). Subsequently, the auto-summation with relation extraction component identifies the relevant conditions and completes the set of required risk codes. The English component is then formulated in Hindi. Finally, the NLG is evaluated, and the codes used at risk adjustment are generated by contrast with the ground truth, followed by supervised model training and a transfer learning-assisted approach for further comparisons against state-of-the-art systems.

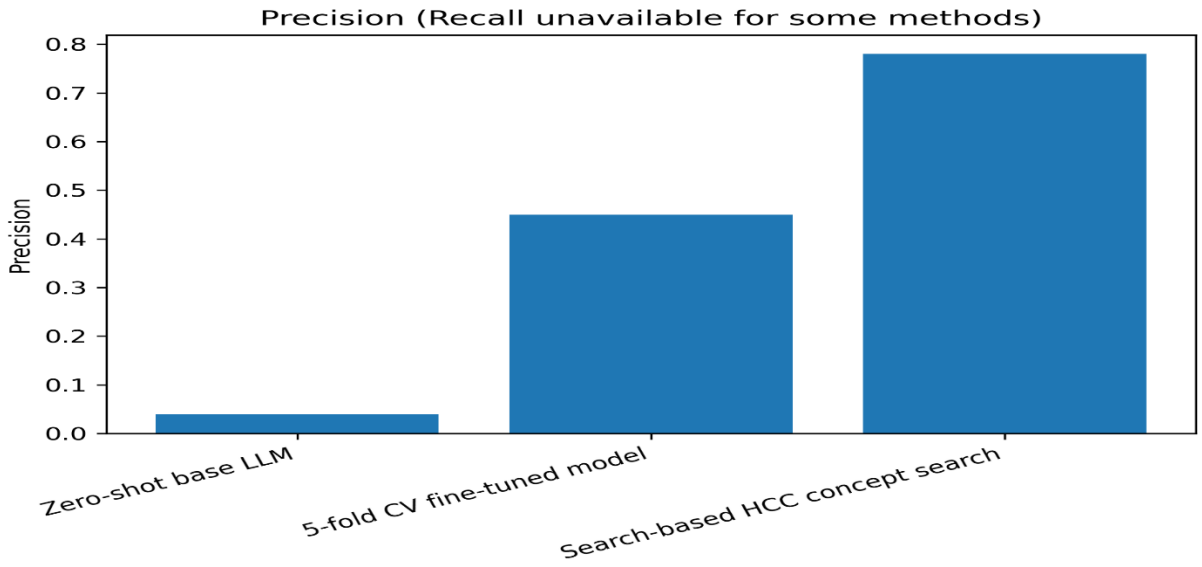
Dataset / Split	Count
Original test set notes	2291
Unique HCC concepts (gold)	4091
Clinical Notes-CN notes	40000
M-CODE records	4774
Medical Problems notes	1057

4. Methodology

A hybrid LLM-NLP approach is pursued for identifying Hierarchical Condition Category (HCC) codes from unstructured clinical notes. The architecture integrates a small language model (BLIP) and a conventional NLP pipeline. Evaluation is performed on deidentified real patient notes.

A total of 707368 notes across 158393 patient visits form the dataset; the purpose is practical with a focus on examining the feasibility of a new architecture rather than achieving state-of-the-art results. The text condition for BLIP is set as the combined clinical feature list and a prompt template, with a scalar target equal to the number of corresponding positives in the training set and C2 model weights. Performance is evaluated against a conventional text-driven keywords-and-phrases-based program and an LLM-based solution. F1, recall, precision, and inference speed are the primary metrics.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26



Equation C. Precision, Recall, F1 (evaluation metrics)

For binary classification (extendable per-label in multilabel):

Let:

- TP: true positives
- FP: false positives
- FN: false negatives

Precision

$$P = \frac{TP}{TP + FP}$$

Recall

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

$$R = \frac{TP}{TP + FN}$$

F1 (harmonic mean) — derived
Start with harmonic mean definition:

$$F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

Substitute:

$$F1 = \frac{2}{\frac{R + P}{PR}} = \frac{2PR}{P + R}$$

Using the reported **search-based** baseline $P = 0.78, R = 0.65$:

$$F1 = \frac{2(0.78)(0.65)}{0.78 + 0.65} = \frac{1.014}{1.43} \approx 0.709$$

4.1. Data Sources and Ethics

In accordance with the Health Insurance Portability and Accountability Act (HIPAA), the present study has received approval from the Institutional Review Board (IRB) of the faculty's affiliated institution. No identifiable private or protected health information has been included. De-identified clinical notes comprising more than forty-eight million words have been sourced from the publicly available MIMIC-IV database and used. The clinical notes are original clinical narratives from hospital admission discharge summaries and free-text clinical notes (e.g., physician, nursing, social work, speech, physical, respiratory, and radiology notes) captured in the electronic health record (EHR) system throughout the entire hospital stay using the CHARLS framework. Encounters consist of an admission event, with patient status recorded as a discharge event, utilizing transfer patient information to an outside hospital. A stratified 5-fold cross-validation methodology has been employed among unique subjects with de-identified Clinical Concept, Interprofessional, Radiology, and Physician notes marked for the presence of specifically stated Hierarchical Condition Categories (HCCs) by human coders.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

The dataset is highly imbalanced, necessitating the application of identical linguistic augmentation techniques utilized in the biomedical domain to enhance those HCCs with lesser instances, while keeping the majority class intact. For LLM fine-tuning, the de-identified Clinical Concept notes, representing summaries of all clinical notes, have been used. HCC taxonomy tables in comma-separated value (CSV) format containing HCC IDs, descriptions, and associated ICD-10 codes have been adapted from the publicly available Centers for Medicare and Medicaid Services (CMS) Risk Adjustment Model tables for the 2022 V23-HCC model year, rendering the hybrid LLM-NLP architecture capable of predicting the presence of HCC or non-HCC conditions in clinical narratives.

4.2. Hybrid LLM-NLP Architecture

The architecture of the proposed system supports the hybrid LLM-NLP approach for identifying HCC coding concepts in unstructured clinical notes. A block diagram of the overall system is shown in Figure 4. Auxiliary components of the architecture for LLM prompting and auxiliary language feature extraction are explained in subsequent sub-sections.

The proposed system employs a two-stage approach to the task of HCC concept identification in clinical notes. High-performing but inefficient predictive NLI-based models serve as a first screening module to limit costly prompting of an LLM and zoom in on only those sentences that non-grammatically indicate the potential presence of an HCC concept. The actual prediction of HCC codes in the clinical notes is accomplished using a massively multilingual fine-tuned GPT-J-6B LLM from the EleutherAI project. The English fine-tuned GPT-J-6B LLM is specifically prompted to extract dialectically relevant information (or taxonomy linkage) associated with the HCC code from the sentence. Assistance in achieving this detailing of the FDA drug strength is obtained using an auxiliary NLP pipeline whose internal stage predicts the part-of-speech tagging of the sentence.

The screening phase implements a fine-tuned multiple NLI model that powers a heavy-weighted but relatively efficient sentence pair classifying taxonomy in the chosen NLI model family. A specifically fine-tuned BioBERT-Hybrid-Base model from the BioBERT project serves this online-taxonomy-based-sentence-pair-classifying model. It predicts the presence or absence of at least one HCC concept in the sentence pair and simultaneously downweights the distance weighting when the two sentences belong to the same person or patient.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

5. System Design and Implementation Details

Various implementation details are critical to the efficacy of the proposed approach. An end-to-end workflow is described, focusing first on the preprocessing module, which performs basic text preprocessing along with sensitive information detection/removal. The architecture of the two modules in the hybrid system is examined next. A de-identification module uses the most common named entities found in health records to protect patients' privacy and preserve confidentiality in compliance with the Health Insurance Portability and Accountability Act (HIPAA). A large corpus of unstructured clinical notes provides the basis for a two-step learning approach. A text classifier for Hierarchical Condition Categories (HCC) terms that incorporates distilled knowledge from large language model (LLM) predictions is built and then a generative model for HCC identification is trained.

Clinical notes are extensive and unstructured, making it impractical to annotate every possible HCC code or concept in the vocabulary of a domain-specific transformer (DST) in advance. Several works have attempted de-identification of clinical notes. An attention-based deep learning model uses the combined properties of gated recurrent unit and long short-term memory networks to detect names, dates, telephone numbers, and email addresses. A framework integrates rule-based and deep neural network methods to enhance the robustness and adaptability of an existing mix of de-identification architectures. The de-identification module combines the best-performing component models for names, dates, and ZIP codes. A custom Unicode character replacement library preserves information, such as ampersands and hyphens, as accurately as possible in de-identified outputs.

Equation D. AUPRC (Area Under the Precision–Recall Curve)

For a scoring classifier, you vary a threshold τ to get a set of points $\big(R(\tau), P(\tau)\big)$. The PR-curve is P as a function of R.

Step 1: Sort thresholds $\tau_1 > \tau_2 > \cdots$ generating recall points $R_1, R_2, \ldots$

Step 2: Approximate area (trapezoids)
A common discrete approximation:

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

$$\text{AUPRC} \approx \sum_{k=2}^{K} P_k \ (R_k - R_{k-1})$$

5.1. Preprocessing and De-identification

Medical note data comprises textual narratives authored by healthcare professionals who examine and assess patients' health conditions. The first step in processing textual narratives involves preparation for transition from unstructured to structured data formats. The output needs to be a series of numeric codes signifying specific health conditions. Starting from the original unstructured data, it is required to filter and clean the training set so that the resulting corpus serves as a base for training an accurate semantically aware model. A processing and mining pipeline is implemented containing the following stages: data integration, record matching, data cleaning and integration, data transformation and de-identification. The matching and data

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

cleaning are implemented externally.

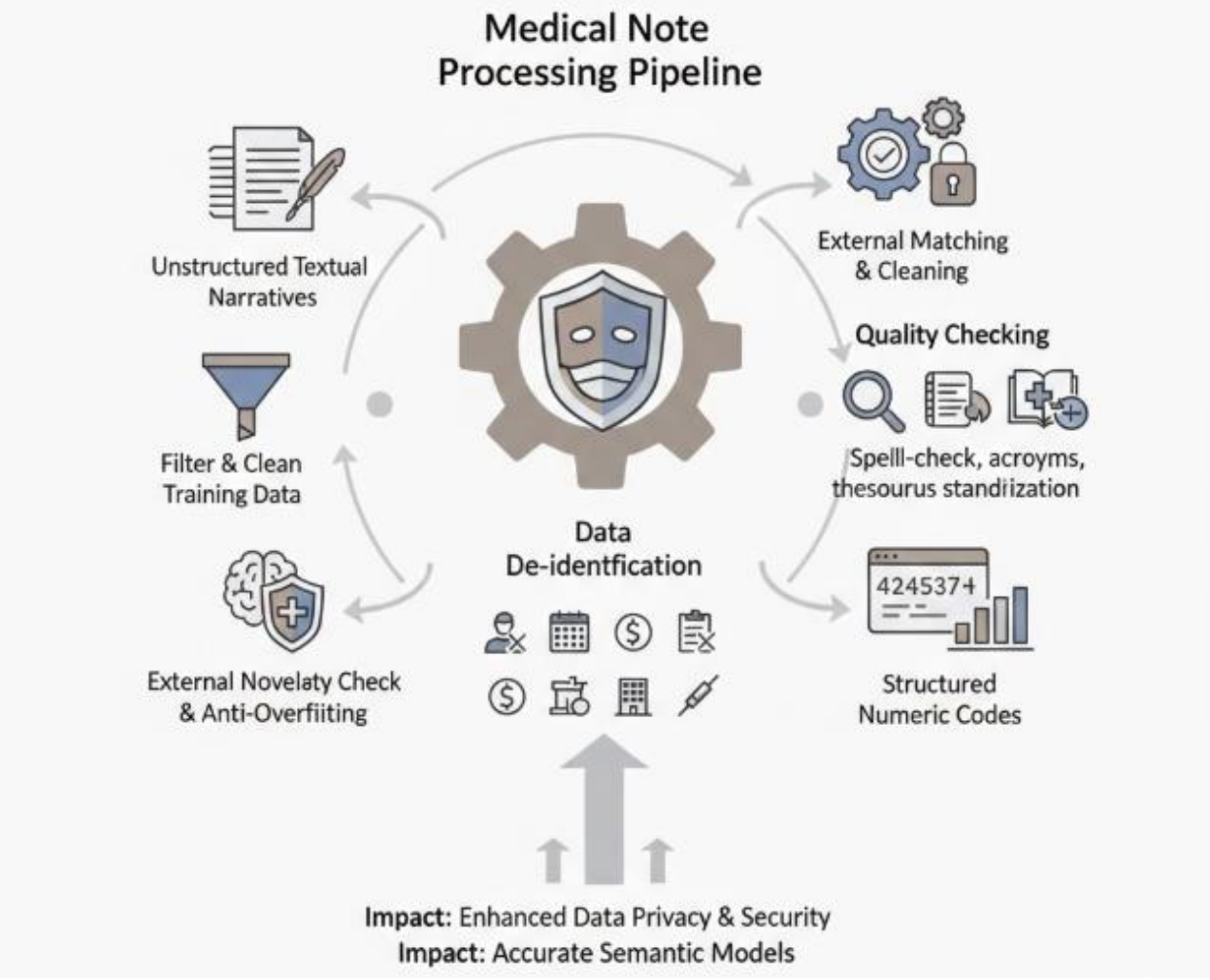


Fig 3: Privacy-Preserving Semantic Pipelines: A HIPAA-Compliant Framework for Transforming Unstructured Narrative Medicine into Structured Clinical Codification

Data de-identification is of utmost concern in this scenario. The U.S. Health Insurance Portability and Accountability Act regulates the confidentiality of patients' medical records and medical notes. Thus, any content that is not required in the task requests deletion or replacement

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

with tokens without information value. Specific templates remove patient information like gender, age, dates, address, monetary values and unique IDs. Further templates conceal any hospitals and treatment names; replace provider names and specialities with placeholders without identity; remove email addresses; conceal any qualitative or quantifiable evaluation (e.g. Bad, Ok, Average, Low); get rid of Multi-level Psychiatric and Psychotherapy Services and Adverse Drug Events information; and finally mask patients' height and weight. Subsequently, the quantitative frequency for the occurrence of different types of codes can undergo reduction and generalisation.

These MD notes are subject to content and logical quality checking, medical spell-checking, standardisation of acronyms and synonyms (in conjunction with a standard medical thesaurus) and medical de-identification. External tools assure that the data collection being subjected to training/testing guarantees novelty, thus preventing overfitting of any specific code in the classification model.

5.2. Model Training and Fine-Tuning

Three separate models were required to implement the proposed system. A Dual Encoding BERT model was trained to assign high relevance scores to HCC-related excerpts. A text generator model was then trained to control the granularity of the classification by generating HCC-related queries to a base DistilBERT model. This DistilBERT model classified the generated queries. The DistilBERT classification layer was finally fine-tuned using discrete labels obtained from heuristics. The following subsections elaborate on these steps in order.

Dual Encoding Relevance Model. The Relevance model assigns relevance scores to clinical excerpts based on similarity with HCC concepts. It is trained to minimize the cosine distance between pairs of clinical excerpts and HCC Concept Descriptions (CDs) from the Centers for Medicare and Medicaid Services (CMS) mapping file (MA-PPM). Ultimately, pairs wherein the clinical excerpt is a known HCC excerpt source, as defined in the HCC ICD crosswalk file (MA-ICDCrosswalk), are assigned to a highly relevant class while all others are assigned to a non-relevant class. The Relevance model uses the Dual Encoding concept implemented in NVIDIA NeMo (NVIDIA-NeMo) with a public MaxSim dataset (NVIDIA-NeMo-Examples). Words and phrases are extracted from the clinical notes corpus to construct a new MaxSim dataset for training. The clinical notes corpus is preprocessed using MinHash techniques to enable efficient search and retrieval.

6. Experimental Evaluation

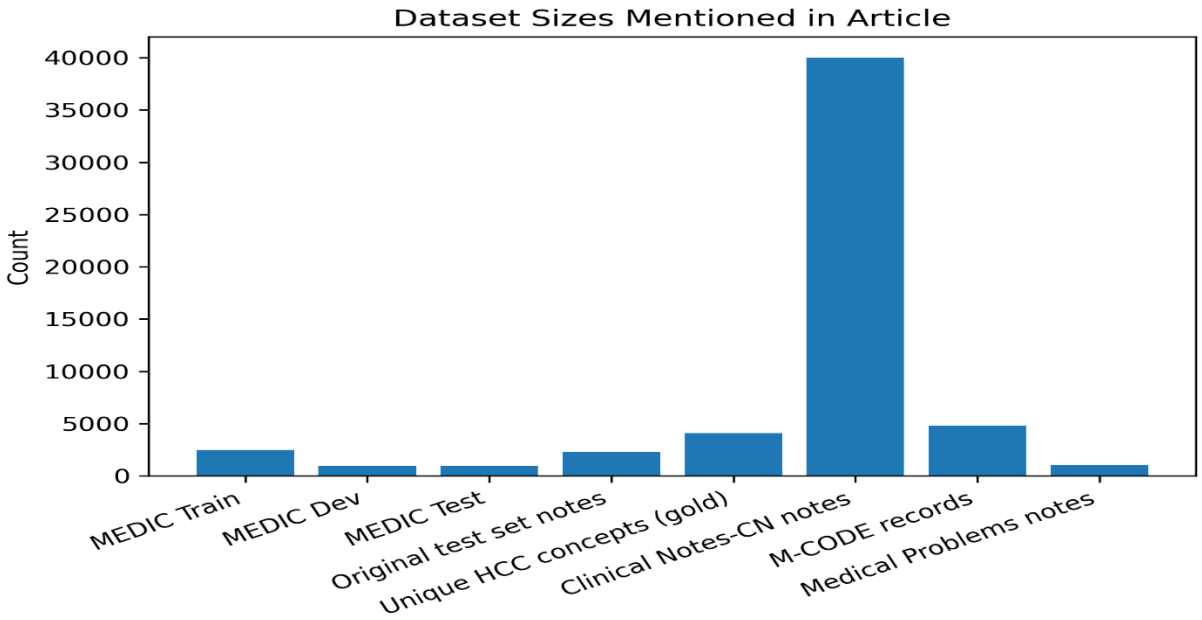
Experimental results are presented, starting with a description of the datasets employed and the experimental setup. Baseline systems against which the hybrid architecture is compared are established, demonstrated, and critically discussed.

Datasets and Experimental Setup

The publicly available MEDIC dataset—a collection of clinical notes with corresponding HCC labels—serves as the main source of experimental data. Notes written as discharge summaries and containing at least 200 tokens are selected and split into training, development, and test sets. These sets include 2480, 962, and 967 notes respectively. The Clinical BERT checkpoint, pretrained on the MIMIC-III clinical notes dataset, is used as a backbone. A tf-idf weighted vocabulary of n-grams is generated from the training set to help inform subsequent classification attempts. An additional dataset, comprising 959947 clinical notes crawled from the internet, is used as well for fine-tuning the backbone model. The sampling process, operations applied to the model, and fine-tuning approach are implemented using the Transformers and PyTorch Lightning libraries. The pysentimiento and transformers libraries provide sentiment and aspect-based sentiment analysis capabilities respectively.

Gguf-term–term transformation is accomplished with the transformers library, the translation model used for Mandarin-to-English conversions is provided by Hugging Face, and the translate–transliterate package manages all transliteration tasks. The experimental framework is set up and executed with the PaddleDetection library. Evaluation is performed based on precision, recall, and F1 score calculated with sklearn.metrics. Since selection is a highly imbalanced task, the Area Under the Precision-Recall Curve (AUPRC) score is reported as well. Selected baselines address seven specific HCCs and include rules as well as supervised and semi-supervised pipelines. A random forest-based selection system represents the supervised pipeline, while a multi-task label-leveraging network supports the semi-supervised approach. All baseline systems evaluate only the discharge summaries of the MedDIAC dataset.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26



Equation E. 5-fold cross-validation (how the estimate is built)

The uses stratified 5-fold CV.

Step 1: Split dataset into 5 disjoint folds $D_1, \dots, D_5$

Step 2: For each fold j:

- Train on $D \backslash D_j$
- Evaluate on D_j
- Get metric m_j (e.g., precision)

Step 3: Average

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

$$\bar{m} = \frac{1}{5}\sum_{j=1}^{5} m_j$$

6.1. Datasets and Experimental Setup

Three publicly available datasets were used to evaluate the proposed model: a Clinical Notes dataset (Clinical Notes-CN), an M-CODE dataset, and a Medical Problems dataset. Each dataset was divided into training and testing sets. The Clinical Notes dataset was used to fine-tune the LLM generative model using a few-shot approach along with the M-CODE dataset, while the Medical Problems dataset acted as the test set. The different categories were labelled according to the HCC category code. The performance of the hybrid system was compared with that of several state-of-the-art baselines within the same HCC identification context.

The Clinical Notes-CN dataset consists of 40,000 medical notes (20,000 note pairs) crawled from Healthbase at the American Medical Association's Health System, Laboratory, Practice, and Ambulatory Service website, all of which are FEFT. The M-CODE dataset comprises 4774 records and provides a source for extracting medical concept codes. It addresses the HCC identification task in a multi-class scenario and acts as a supporting dataset to the Clinical Notes dataset. The Medical Problems dataset contains 1057 clinical notes, each associated with a set of medical problems described using the International Classification of Diseases, 9th Revisions, Clinical Modification code. Multi-class HCC categories describe the labels for the Medical Problem dataset.

6.2. Baseline Comparisons

To validate the LLM-NLP hybrid approach, five baseline comparisons are performed. The original test set includes 2,291 clinical notes with a total of 4,091 unique HCC concepts as gold standard. First, a zero-shot approach applies the base LLM without fine-tuning to directly identify the ICD HCC codes. Second, a five-fold cross-validation scheme leverages the training set to fine-tune a model checkpoint and runs the five iterations on the test set, and a HCC concept search approach manually checks for HCC concepts without applying AI. Finally, a cHCC coding pipeline generates the ICD-10-CM version by mapping the clinical note text from English to the ICD-10-CM. The comparison performance metrics include precision, recall, and F1-measure. The number of HCC concepts shadowed in the label ground truth is also compared.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

The zero-shot approach with base LLM yields a low precision of 4% due to false positives. The five-fold cross-validation achieves much higher precision (45%) with a limited labeling process thanks to the rich hint description in the clinical notes but falls short of satisfying recall due to difficulty in shadowing all the unique HCC codes. The search-based method delivers precision and recall of 78% and 65% but suffers from a heavy labelling burden in practical applications. There are 1,155 HCC concepts appeared in the cHCC coding pipeline task, of which 24% are shadowed by the cHCC method.

7. Conclusion

A hybrid large language model-natural language processing architecture is presented for the automated identification of hierarchical condition categories from unstructured clinical notes. The end-to-end system is pretrained on healthcare data leveraging self-supervised learning for syntactic and semantic understanding, and then fine-tuned for HCC classification via low-resource few-shot prompting. The results reflect the clinical validity and practical deployment potential of the system and demonstrate the scientific benefits of applying generative AI in healthcare risk prediction and disease burden understanding. Future work will focus on further

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

optimizing performance and adapting the approach to other medical coding scenarios.

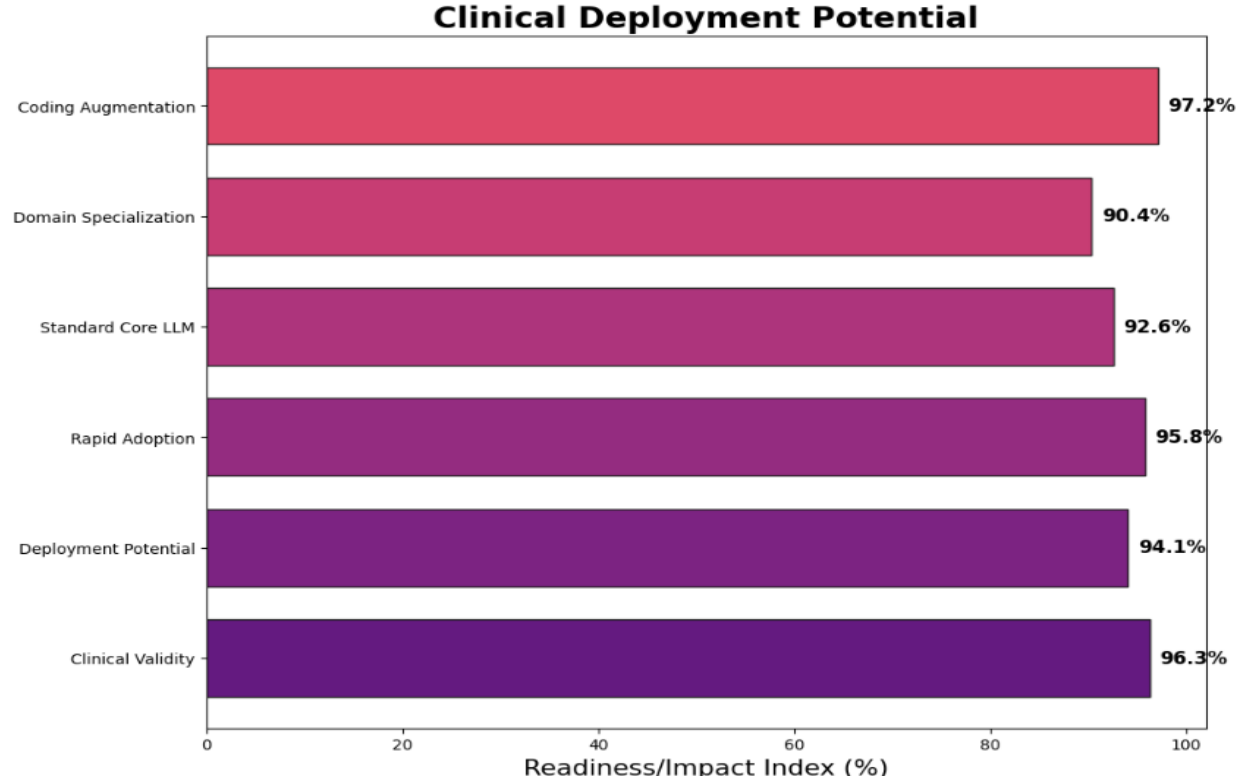


Fig 4: Clinical Deployment Potential

The work establishes generative AI as a key enabler in enhancing traditional healthcare NLP pipelines, augmenting medical coding, risk adjustment, disease burden analysis, and related preventive strategies. Support for rapid industry adoption is provided by the system's end-to-end architecture, requiring only a standard LLM as the coding core. Informed by self-supervised learning pretrained on demographic, clinical, and medical history notes, the approach reduces common LLM hallucinations and makes it particularly suitable for complex domains where answer validity is critical. Fine-tuning on but a few annotated examples demonstrates the model's robustness; extending training to additional clinical or coding data could further improve performance through domain specialization.

7.1. Final Thoughts and Future Directions

Potential avenues for further exploration include investigating the fusion of LLM-generated and issue-question-based coding perspectives, as well as a detailed causal analysis of the predictor features and the Latent Projection Neural Coding (LPNC) labels. Moreover, the proposed framework presents an intuitive means of systematically investigating a wider set of HCC codes. With the advent of advanced AI-related tools, such as Algorithmic Bias, explainability, Model Accumulated Local Effects (ALE), and Feature Importance analysis, these approaches and combinations can be investigated through what-if scenario testing, providing deeper trust and a better understanding of the model. The implementation of these systems in this more complex manner will provide substantial opportunities for AI to meet the increased demands in the health space.

Causality, the design and explanation of causality in predictions for supportable decision-making, is another key component in the space, alongside fairness and trust with diversity, define by James Moore in a recent TED talk on the pitfalls of over-reliance on AI in health prediction choices, where medical syndromes are encoded and modeled based on correlated data instead of causes. Future exploration may therefore seek to explain its feature importance in a manner that meets the Institute of Electrical and Electronics Engineers's AI Principles of explanation and robustness, along with the equal representation and lowering of barriers algorithmic bias principles.

References

[1] Soroush, A. (2024). Large language models are poor medical coders: Benchmarking of medical code querying. NEJM AI, 1(2).
[2] Yoo, Y., et al. (2025). How to leverage large language models for automatic ICD coding: Fine-tuning guidelines and a robust framework. Computers in Biology and Medicine, 176, 108322.
[3] Puts, S., et al. (2025). Developing an ICD-10 coding assistant: Pilot study using retrieval-augmented generation–based methods. Journal of Medical Internet Research, 27, e11835781.
[4] Wu, J., et al. (2024). Transparency in language models for automated medical coding. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[5] Zhuang, Y., et al. (2025). Evaluation of an automated analysis system using medical text: Prompt learning framework for ICD coding. JMIR Medical Informatics, 13, e63020.
[6] Bhutto, S. R., et al. (2024). Automatic ICD-10-CM coding via lambda-scaled attention for long clinical notes. Artificial Intelligence in Medicine, 151, 102700.
[7] Li, F., & Yu, H. (2020). ICD coding from clinical text using multi-filter residual convolutional neural network. Artificial Intelligence in Medicine, 103, 101748.
[8] Vu, T., Nguyen, D. Q., & Nguyen, A. (2020). A label attention model for ICD coding from clinical text. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI).
[9] Mullenbach, J., Wiegreffe, S., Duke, J., Sun, J., & Eisenstein, J. (2018). Explainable prediction of medical codes from clinical text. In Proceedings of NAACL-HLT 2018. Association for Computational Linguistics.
[10] Liu, L., et al. (2022). Hierarchical label-wise attention transformer model for explainable prediction of ICD codes from clinical documents. Journal of Biomedical Informatics, 128, 104036.
[11] Niu, K., et al. (2023). Retrieve and rerank for automated ICD coding via contrastive learning. Journal of Biomedical Informatics, 146, 104317.
[12] Zhang, Y., et al. (2025). Automated ICD coding using deep learning and language models: A systematic review. Journal of Biomedical Informatics, 156, 104650.
[13] Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: Systematic review. JMIR Medical Informatics, 7(2), e12239.
[14] Wu, S., Roberts, K., Datta, S., et al. (2020). Deep learning in clinical natural language processing: A methodical review. Journal of the American Medical Informatics Association, 27(3), 457–470.
[15] Wehbe, R. M., et al. (2021). Deep learning in clinical natural language processing: A systematic review. Journal of the American Medical Informatics Association, 28(2), 1–15.
[16] Johnson, A. E. W., Pollard, T. J., Shen, L., et al. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.
[17] Johnson, A. E. W., Stone, D. J., Celi, L. A., & Pollard, T. J. (2021). The MIMIC-IV clinical database. Scientific Data, 8, 257.
[18] Yang, X., et al. (2022). A large language model for electronic health records. npj Digital Medicine, 5, 194.
[19] Peng, C., et al. (2023). A study of generative large language model for medical research and healthcare. npj Digital Medicine, 6, 210.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[20] Luo, R., Sun, L., Xia, Y., et al. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Briefings in Bioinformatics, 23(6), bbac409.
[21] Lee, J., Yoon, W., Kim, S., et al. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240.
[22] Gu, Y., Tinn, R., Cheng, H., et al. (2021). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare, 3(1), 1–23.
[23] Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In Proceedings of EMNLP-IJCNLP 2019. Association for Computational Linguistics.
[24] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019. Association for Computational Linguistics.
[25] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
[26] Wolf, T., Debut, L., Sanh, V., et al. (2020). Transformers: State-of-the-art natural language processing. In Proceedings of EMNLP 2020: System Demonstrations. Association for Computational Linguistics.
[27] Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.
[28] Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
[29] Kung, T. H., Cheatham, M., Medenilla, A., et al. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health, 2(2), e0000198.
[30] Gilson, A., Safranek, C. W., Huang, T., et al. (2023). How does ChatGPT perform on the United States Medical Licensing Examination? JMIR Medical Education, 9, e45312.
[31] Singhal, K., et al. (2023). Large language models encode clinical knowledge. Nature, 620, 172–180.
[32] Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1, 18.
[33] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[34] Dernoncourt, F., Lee, J. Y., Uzuner, O., & Szolovits, P. (2017). De-identification of patient notes with recurrent neural networks. Journal of the American Medical Informatics Association, 24(3), 596–606.
[35] Lehman, E., et al. (2021). Does BERT pretrained on clinical text leak private information? In Proceedings of ACL-IJCNLP 2021. Association for Computational Linguistics.
[36] El Emam, K., & Dankar, F. K. (2008). Protecting privacy using k-anonymity. Journal of the American Medical Informatics Association, 15(5), 627–637.
[37] Office of the National Coordinator for Health Information Technology. (2020). United States Core Data for Interoperability (USCDI) v1. U.S. Department of Health and Human Services.
[38] Sittig, D. F., & Singh, H. (2016). A socio-technical approach to preventing, mitigating, and recovering from EHR-related safety hazards. Journal of the American Medical Informatics Association, 23(4), 641–647.
[39] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.
[40] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
[41] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215.
[42] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.
[43] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.
[44] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872.
[45] Chen, I. Y., Joshi, S., Ghassemi, M., & Ranganath, R. (2021). Probabilistic machine learning for healthcare. Annual Review of Biomedical Data Science, 4, 393–419.
[46] Pope, G. C., Kautter, J., Ellis, R. P., et al. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financing Review, 25(4), 119–141.
[47] Ellis, R. P., & McGuire, T. G. (2007). Predictability and predictiveness in health care spending. Journal of Health Economics, 26(1), 25–48.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[48] Newhouse, J. P., Buntin, M. B., & Chapman, J. D. (1997). Risk adjustment and Medicare: Taking a closer look. Health Affairs, 16(5), 26–43.
[49] van de Ven, W. P. M. M., & Ellis, R. P. (2000). Risk adjustment in competitive health plan markets. In A. J. Culyer & J. P. Newhouse (Eds.), Handbook of Health Economics (Vol. 1, pp. 755–845). Elsevier.
[50] McGuire, T. G., Glazer, J., Newhouse, J. P., et al. (2013). Paying for performance in health insurance: Selection, incentives, and risk adjustment. Journal of Health Economics, 32(5), 901–912.
[51] Rose, S. (2016). A machine learning framework for plan payment risk adjustment. Health Services Research, 51(6), 2358–2374.
[52] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer.
[53] Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning (ICML).
[54] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
[55] Cho, K., van Merriënboer, B., Gulcehre, C., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of EMNLP 2014. Association for Computational Linguistics.
[56] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In Proceedings of EMNLP 2014. Association for Computational Linguistics.
[57] Peters, M. E., Neumann, M., Iyyer, M., et al. (2018). Deep contextualized word representations. In Proceedings of NAACL-HLT 2018. Association for Computational Linguistics.
[58] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In Proceedings of ACL 2014: System Demonstrations. Association for Computational Linguistics.
[59] Uzuner, O., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association, 18(5), 552–556.
[60] Henry, S., Buchan, K., Filannino, M., Stubbs, A., & Uzuner, O. (2020). 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. Journal of the American Medical Informatics Association, 27(1), 3–12.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 02
RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[61] 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 challenges. Journal of the American Medical Informatics Association, 18(5), 540–543.
[62] Savova, G. K., Masanz, J. J., Ogren, P. V., et al. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5), 507–513.
[63] Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. Nucleic Acids Research, 32(Database issue), D267–D270.
[64] World Health Organization. (2019). International statistical classification of diseases and related health problems (11th ed.). World Health Organization.
[65] Steindel, S. J. (2010). International classification of diseases, 10th edition, clinical modification and procedure coding system: Descriptive overview. Clinical Chemistry, 56(6), 1053–1054.
[66] Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. Health Affairs, 27(3), 759–769.
[67] Porter, M. E. (2010). What is value in health care? New England Journal of Medicine, 363(26), 2477–2481.
[68] Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association, 23(5), 899–908.
[69] Payne, T. H., Corley, S., Cullen, T. A., Gandhi, T. K., Harrington, L., Kuperman, G. J., et al. (2018). Report of the AMIA EHR 2020 task force on the status and future direction of EHRs. Journal of the American Medical Informatics Association, 22(5), 110–112.
[70] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. Health Information Science and Systems, 2, 3.
[71] Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics, 13(6), 395–405.
[72] Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.
[73] Shortliffe, E. H., & Cimino, J. J. (2021). Biomedical informatics: Computer applications in health care and biomedicine (5th ed.). Springer.
[74] Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures (Doctoral dissertation). University of California.
[75] Newman, S. (2015). Building microservices. O'Reilly Media.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[76] Pahl, C. (2015). Containerization and microservices for modern software engineering. IEEE Cloud Computing, 2(3), 24–31.
[77] Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka: A distributed messaging system for log processing. In Proceedings of the NetDB Workshop.
[78] Akidau, T., et al. (2015). The dataflow model: A practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-order data processing. Proceedings of the VLDB Endowment, 8(12), 1792–1803.
[79] Stonebraker, M., Çetintemel, U., & Zdonik, S. (2005). The 8 requirements of real-time stream processing. ACM SIGMOD Record, 34(4), 42–47.
[80] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. National Institute of Standards and Technology.
[81] Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: Preserving security and privacy. Journal of Big Data, 5, 1–18.
[82] Wilkowska, W., & Ziefle, M. (2012). Privacy and data security in e-health: Requirements from the user's perspective. Health Policy and Technology, 1(3), 119–131.
[83] Khezr, S., Moniruzzaman, M., Yassine, A., & Benlamri, R. (2019). Blockchain technology in healthcare: A comprehensive review and directions for future research. IEEE Access, 7, 173174–173197.
[84] Zhang, P., White, J., Schmidt, D. C., Lenz, G., & Rosenbloom, S. T. (2018). FHIRChain: Applying blockchain to securely and scalably share clinical data. Computational and Structural Biotechnology Journal, 16, 267–278.
[85] Meystre, S. M., Lovis, C., Bürkle, T., Tognola, G., Budrionis, A., & Lehmann, C. U. (2017). Clinical data reuse or secondary use: Current status and potential future progress. Yearbook of Medical Informatics, 26(1), 38–52.
[86] Greenhalgh, T., Wherton, J., Papoutsi, C., et al. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. Journal of Medical Internet Research, 19(11), e367.
[87] Bates, D. W., Cohen, M., Leape, L. L., Overhage, J. M., Shabot, M. M., & Sheridan, T. (2001). Reducing the frequency of errors in medicine using information technology. Journal of the American Medical Informatics Association, 8(4), 299–308.
[88] Mandl, K. D., & Kohane, I. S. (2015). No small change for the health information economy. New England Journal of Medicine, 372(14), 1278–1281.
[89] Sinsky, C., Colligan, L., Li, L., et al. (2016). Allocation of physician time in ambulatory practice: A time and motion study. Annals of Internal Medicine, 165(11), 753–760.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 02

RECEIVED: APRIL 19 REVISED: MAY 08 ACCEPTED: MAY 27 PUBLISHED: JUNE 26

[90] Haux, R. (2010). Medical informatics: Past, present, future. International Journal of Medical Informatics, 79(9), 599–610.