Medical Diagnosis Coding Automation:

Similarity Search vs. Generative AI

Vanessa Klotzman MSc[1,2]

[1]Department of Informatics , University of California, Irvine, 6210 Donald Bren Hall, Irvine, 92697-3425, CA, United States.
[2]Research Institute, Children's Hospital of Orange County, 1201 W. La Veta Ave., Orange County, 92868, CA, United States.

Contributing authors: vklotzma@uci.edu;

Abstract

Objective: This study aims to predict ICD-10-CM codes for medical diagnoses from short diagnosis descriptions and compare two distinct approaches: similarity search and using a generative model with few-shot learning.
Materials and Methods: The text-embedding-ada-002 model was used to embed textual descriptions of 2023 ICD-10-CM diagnosis codes, provided by the Centers provided for Medicare & Medicaid Services. GPT-4 used few-shot learning. Both models underwent performance testing on 666 data points from the eICU Collaborative Research Database.
Results: The text-embedding-ada-002 model successfully identified the relevant code from a set of similar codes 80% of the time, while GPT-4 achieved a 50 % accuracy in predicting the correct code.
Discussion: The work implies that text-embedding-ada-002 could automate medical coding better than GPT-4, highlighting potential limitations of generative language models for complicated tasks like this.
Conclusion: The research shows that text-embedding-ada-002 outperforms GPT-4 in medical coding, highlighting embedding models' usefulness in the domain of medical coding.

Keywords: Embeddings, Large Language Models, ICD Codes, Automated Medical Coding

1 Background and Significance

The International Classification of Disease (ICD), established by the World Health Organization (WHO) is the universally recognized and standardized system for medical coding worldwide. It provides a comprehensive framework for categorizing diseases, health conditions, and related information, facilitating accurate and consistent documentation, data sharing, and research across the global healthcare community. It is employed by healthcare providers worldwide to categorize diseases and conditions. Medical coding involves the assignment of ICD codes like ICD-10-CM codes to classify diagnoses and reasons for visits in all healthcare settings, is essential for guiding clinical decisions, tracking diseases, and impacting healthcare financing[1, 2].

Medical coding is traditionally manual, with coders translating physicians' notes into the appropriate ICD codes while adhering to complex guidelines. In this process, highly trained medical coders assign ICD (International Classification of Diseases) codes to patient encounters based on the information found in clinicians' notes, however, manual ICD coding is time-consuming and error-prone, making the quality and productivity of coding a matter of concern in practice. The process is error-prone [3–9] due to the complexity of medical language and coding guidelines. Coders often need help with subtle differences between disease subtypes, leading to misclassification. Physicians' use of abbreviations and synonyms which adds to the ambiguity [10, 11]. Making this a non-trivial task for humans. Furthermore, inexperienced coders may incorrectly assign separate codes to related diagnoses, a problem called unbundling, which can result in costly mistakes [12]. These coding inaccuracies have substantial financial implications, contributing to an estimated annual expenditure of $25 billion in the United States, as reported by Lang et al. [13] Farkas et al. [14]. With recent AI technologies (e.g., NLP), automated medical coding has the potential to support clinical coders better.

Automated Medical Coding (AMC) is the idea that artificial intelligence can automate clinical coding. In recent years, there has been a significant increase in AMC-related work [14–32] through deep learning. Although research in this field has grown, this problem is far from being solved [18, 33].For instance, automated coding remains a complex problem because extracting knowledge from patients' clinical records is challenging. These records are not uniformly structured, the medical field's terminologies can be complicated for non-professionals to comprehend, and physicians often have different ways of describing symptoms, leading to various descriptions for the same disease.

Embeddings [34] in Natural Language Processing (NLP) represent words as real-valued vectors. These vectors can capture the meaning of words in such a way that words closer together in the vector space are expected to have similar meanings. In clinical NLP [35], embeddings are helpful for analyzing medical data and texts, aiding decision-making and research. The use of word embeddings in Automated Medical Coding (AMC) systems [36–43] is increasingly being explored as it has the potential to bridge the gap between the informal language of medical diagnoses and the formal language of ICD code descriptions. For instance CAIC [27] uses cross-textual attention to match parts of medical notes with ICD codes. While GatedCNN-NCI [18] creates a network linking every aspect of medical notes to ICD codes. BiCapsNetLE [44] integrates ICD code descriptions into word embeddings of clinical notes, enhancing alignment. DLAC [45] employs a description-based label attention mechanism, focusing on the correlation between the descriptions of ICD codes and the features of medical notes. ICDBigBird uses a Graph Convolutional Network (GCN) and ehances the ICD code emebddings by using their relational structure.

Even though, there is a growing body of work for utilizing embeddings in clinical coding, there has been a growing interest of what a Large Language Model(LLM) can do in the health sector due to their ability of understanding, generating, and predicting new content.

As the interest in Large Language Models (LLMs) continues to grow in the health sector, as evidenced by multiple recent studies studies [46–57], our objective is to compare the effectiveness of two distinct approaches to predict ICD-10-CM codes accurately. We will compare the effectiveness of similarity search, for which we will be using text-embedding-ada-002 [58], and an LLM, in which we will be using GPT-4 [59] from OpenAI to predict ICD-10-CM codes.

2 Materials and Methods

2.1 Data collection

We utilized the diagnosis strings (patient diagnoses) from the eICU Collaborative Research Database [60], which contains data from different critical care units (CCUs) across the United States from patients who were admitted between 2014 and 2015. We selected a subset of 666 patients from the total dataset of 2,710,672 patients. This sample size represents a 99% confidence level with a 5% margin of error [61]. We utilize each patient's current diagnoses from the data we collected, which comprise of the diagnosis string and the corresponding ICD-10 CM codes. The diagnosis strings will serve as inputs to the models, with the ICD-10-CM codes as the outputs. The ICD-10-CM codes will be used for comparison to assess the model's accuracy in prediction. The diagnosis strings in the eICU database are organized in a tiered system. For example, "neurologic—trauma - CNS—intracranial injury—with subarachnoid hemorrhage" shows this: it starts with a general category "neurologic", goes into a more specific "trauma - CNS", then to "intracranial injury", and ends with a detailed aspect "with subarachnoid hemorrhage". Each part of the string represents a deeper level of diagnosis detail. Table 1, shows sample diagnosis strings and their corresponding ICD-10 CM codes from the dataset we are using.

Table 1 eICU Sample Data: Diagnosis & ICD-10 Code

Diagnosis	ICD-10 CM Code
burns/trauma|dermatology|cellulitis	L03.90
burns/trauma|trauma - chest|lung trauma	S27.30
hematology|coagulation disorders|DVT	I80.9

2.2 Model Selection

We utilize OpenAI's text-embedding-ada-002 model [1], as it surpasses previous models in text search and text similarity from OpenAI. We evaluate the effectiveness of the embedding model relative to the latest version of GPT-4, which we operated using the Microsoft OpenAI Azure Service. Our selection of only the text-embedding-ada-002 model and GPT-4 was due to the constraints set by the terms and conditions of using the PhysioNet dataset [2].

2.3 text-embedding-ada-002

In this work, we utilized the text-embedding-ada-002 model to embed the textual descriptions of 2023 ICD-10-CM diagnosis codes [3] , as provided by the Centers for Medicare & Medicaid Services source. After generating these embeddings, our primary objective was to evaluate their performance. To do so, we used the dataset of diagnosis strings obtained from the eICU dataset.

To assess the accuracy of matching medical diagnoses (diagnosis strings) with their respective ICD-10-CM codes, we inputted these diagnosis strings into the text-embedding-ada-002 model. Our objective was to determine if this single model could accurately return the closest ICD-10-CM code based on the ICD-10 description. Additionally, the model returned the top four ICD-10-CM codes for each medical condition. We selected four as the default value for similarity searches by vector, because this is specified as the standard setting in the LangChain documentation [4]. The workflow of

[1] https://openai.com/blog/new-and-improved-embedding-model
[2] https://physionet.org/news/post/415
[3] https://www.cms.gov/medicare/coding-billing/icd-10-codes
[4] https://api.python.langchain.com/

how the embeddings function is clearly illustrated in Figure 1 The embeddings are generated and then stored in the vector database. These embeddings correspond to the textual descriptions of the 2023 ICD-10-CM diagnosis codes. When a user submits a query as a medical diagnosis (referred to as the diagnosis string), the system searches the database for embeddings similar to the embedding of the query. Finally, the system retrieves the closest ICD-10-CM codes based on the similarity between embeddings, providing relevant matches for the medical diagnosis.

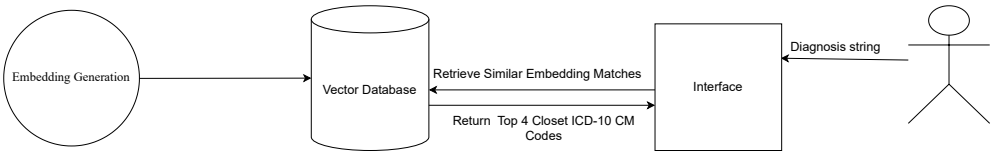

Fig. 1 Visualizing Automated ICD-10 Code Prediction Process: Streamlining Medical Coding

2.4 GPT-4

We prompted GPT-4 with few-shot prompting to assess its capability in medical coding. Few shot prompting [62] was selected because large language models have notable zero-shot abilities, but they tend to perform poorly in complex tasks when using zero-shot settings. Few-shot prompting serves as a method to enable in-context learning, where demonstrations in the prompt help direct the model toward better performance. Figure 2 contains the prompt we used. The examples for the prompt were acquired by clustering the diagnosis strings. We used K-means clustering [63] to group our diagnosis strings and found that 8 clusters worked best. The ideal number of eight clusters was determined using the elbow method [64], which evaluates the within-cluster sum of squares across a range of 1 to 20 possible clusters; the 'elbow' point, where there is a significant decrease in within-cluster dissimilarity, indicates the most suitable number of clusters. We picked a range between 1 and 20 as it is a manageable number of clusters that can be effectively interpreted and analyzed. From

each cluster, we selected the diagnosis closest to the cluster's center in the vector space as the most representative example of the cluster; these representative examples were then used in the few-shot prompt.

To optimize GPT-4 for medical coding, we experimented with different temperature settings (0.1, 0.5, 0.9) on the sample of 666 diagnoses collected for this study, representing a 99% confidence interval and a 5% margin of error. Our results showed that a temperature of 0.1 was effective, as it balanced the model's creative outputs and the need for accurate, deterministic responses in medical coding.

```
Role: Medical Coder
Objective:
Your task is to accurately assign the correct ICD-10-CM code for each
patient's condition based on their medical diagnosis.
Examples:
Description: neurologic|disorders of vasculature|stroke
Output: I67.8
Description: infectious diseases|systemic/other infections|sepsis
Output: A41.9
Description: pulmonary|disorders of vasculature|pulmonary embolism
Output: I26.99
Description: burns/trauma|trauma - CNS|intracranial injury
Output: S06.9
Description: cardiovascular|ventricular disorders|congestive heart failure
Output: I50.9
Description: gastrointestinal|GI bleeding / PUD|peptic ulcer disease
Output: K27.9
Description: pulmonary|respiratory failure|acute respiratory failure
Output: J96.00
Description: renal|disorder of kidney|acute renal failure
Output: N17.9
Description: ${input_text}$
Output:
```

Fig. 2 GPT-4 Prompt

3 Results

The text-embedding-ada-002 model achieved an 80% accuracy rate in identifying the correct ICD-10-CM codes from the retrieved similar codes, outperforming GPT-4,

which achieved a 50% accuracy rate in the same task. This suggests that embedding models, like text-embedding-ada-002, can offer improved accuracy and efficiency in medical coding.

4 Discussion

In this study, we discovered that embedding models like text-embedding-ada-002 could potentially be more effective than GPT-4, a large language model. The critical advantage of embeddings lies in their focus on the semantic similarity of words, an aspect vital for accurately matching medical diagnoses with ICD codes, as this technique allows for a more precise understanding and interpretation of medical terminology, which is crucial in medical coding. Embeddings analyze the context and meaning of words more concentratedly, leading to higher accuracy in identifying relevant codes.

Moreover, when assessing the feasibility of using embedding models like text-embedding-ada-002, it becomes evident that these models align well with medical coding requirements. They offer a more focused approach, potentially assisting in accurately linking diagnoses with the correct ICD codes, which demands precision. It suggests that embedding models better fit medical coding tasks compared to more generative models like GPT-4, which handle a broader range of data.

In contrast, GPT-4 processes a wide range of data and contexts. While this versatility is helpful for general tasks, it can lead to less precision in specialized areas like medical coding, where specific terminology and accurate coding are essential. GPT-4's handling of vast information might make it more challenging to differentiate between similar medical terms and codes, potentially affecting its performance in this field.

5 Conclusion

The results indicate that embedding models like text-embedding-ada-002 appear more suitable for medical coding tasks than large language models like GPT-4. This

result could be primarily due to text-embedding-ada-002's focused approach on the semantic similarity of words, which has led to an 80% accuracy in identifying ICD-10-CM codes, significantly surpassing GPT-4's 50% accuracy. Embedding models like text-embedding-ada-002 are a more practical choice for medical coding due to their precision in analyzing and understanding medical terminology. On the other hand, GPT-4, although capable of broad data processing, may prove less effective in specialized fields such as medical coding, where accuracy and specific terminology are crucial. Hence, for precision-dependent tasks such as medical coding, embedding models like text-embedding-ada-002 could offer a more suitable solution than generative models like GPT-4.

References

[1] Aalseth P. Medical Coding: What it is and how it Works. Jones & Bartlett Publishers; 2014.

[2] Borman KR. Medical Coding in the United States: Introduction and Historical Overview. Principles of Coding and Reimbursement for Surgeons. 2017;p. 3–11.

[3] O'malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health services research. 2005;40(5p2):1620–1639.

[4] Champagnie SJ. Medicare Loses Billions to Billing Errors. In: Proceedings of the Ninth International Conference on Engaged Management Scholarship (2019); 2019. .

[5] Asadi F, Hosseini MA, Gomar T, Sabahi A. Factors Affecting Clinical Coding Errors. Shiraz E-Medical Journal. 2022;23(9).

[6] Lloyd SS, Rissing JP. Physician and coding errors in patient records. Jama. 1985;254(10):1330–1336.

[7] Cipparone CW, Withiam-Leitch M, Kimminau KS, Fox CH, Singh R, Kahn L. Inaccuracy of ICD-9 codes for chronic kidney disease: a study from two practice-based research networks (PBRNs). The Journal of the American Board of Family Medicine. 2015;28(5):678–682.

[8] Benesch C, Witter D, Wilder A, Duncan P, Samsa G, Matchar D. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology. 1997;49(3):660–664.

[9] Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. Journal of the American Medical Informatics Association. 2016;23(e1):e20–e27.

[10] Xie P, Xing E. A neural architecture for automated ICD coding. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2018. p. 1066–1076.

[11] Sheppard JE, Weidner LC, Zakai S, Fountain-Polley S, Williams J. Ambiguous abbreviations: an audit of abbreviations in paediatric note keeping. Archives of disease in childhood. 2008;93(3):204–206.

[12] Alonso V, Santos JV, Pinto M, Ferreira J, Lema I, Lopes F, et al. Problems and barriers during the process of clinical coding: a focus group study of coders' perceptions. Journal of medical systems. 2020;44:1–8.

[13] Lang D. Consultant report-natural language processing in the health care industry. Cincinnati Children's Hospital Medical Center, Winter. 2007;6.

[14] Farkas R, Szarvas G. Automatic construction of rule-based ICD-9-CM coding systems. In: BMC bioinformatics. vol. 9. Springer; 2008. p. 1–9.

[15] Teng F, Liu Y, Li T, Zhang Y, Li S, Zhao Y. A review on deep neural networks for ICD coding. IEEE Transactions on Knowledge and Data Engineering. 2022;35(5):4357–4375.

[16] Diao X, Huo Y, Zhao S, Yuan J, Cui M, Wang Y, et al. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. International journal of medical informatics. 2021;153:104543.

[17] Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR. A systematic literature review of automated clinical coding and classification systems. Journal of the American Medical Informatics Association. 2010;17(6):646–651.

[18] Ji S, Sun W, Dong H, Wu H, Marttinen P. A unified review of deep learning for automated medical coding. arXiv preprint arXiv:220102797. 2022;.

[19] Chen Y, Chen H, Lu X, Duan H, He S, An J. Automatic ICD-10 coding: Deep semantic matching based on analogical reasoning. Heliyon. 2023;9(4).

[20] Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. npj Digital Medicine. 2023;6(1):16.

[21] Catling F, Spithourakis GP, Riedel S. Towards automated clinical coding. International journal of medical informatics. 2018;120:50–61.

[22] Edin J, Junge A, Havtorn JD, Borgholt L, Maistro M, Ruotsalo T, et al. Automated Medical Coding on MIMIC-III and MIMIC-IV: A Critical Review and Replicability Study. arXiv preprint arXiv:230410909. 2023;.

[23] Moons E, Khanna A, Akkasi A, Moens MF. A comparison of deep learning methods for ICD coding of clinical records. Applied Sciences. 2020;10(15):5262.

[24] Ramalho A, Souza J, Freitas A. The use of artificial intelligence for clinical coding automation: a bibliometric analysis. In: International Symposium on Distributed Computing and Artificial Intelligence. Springer; 2020. p. 274–283.

[25] Jones KA, Beecroft NJ, Patterson ES. Towards computer-assisted coding: A case study of 'charge by documentation'software at an endoscopy clinic. Health Policy and Technology. 2014;3(3):208–214.

[26] Sonabend A, Cai W, Ahuja Y, Ananthakrishnan A, Xia Z, Yu S, et al. Automated ICD coding via unsupervised knowledge integration (UNITE). International journal of medical informatics. 2020;139:104135.

[27] Teng F, Ma Z, Chen J, Xiao M, Huang L. Automatic medical code assignment via deep learning approach for intelligent healthcare. IEEE journal of biomedical and health informatics. 2020;24(9):2506–2515.

[28] Kaur R, Ginige JA, Obst O. AI-based ICD coding and classification approaches using discharge summaries: A systematic literature review. Expert Systems with Applications. 2023;213:118997.

[29] Boytcheva S. Automatic matching of ICD-10 codes to diagnoses in discharge letters. In: Proceedings of the second workshop on biomedical natural language processing; 2011. p. 11–18.

[30] Perotte A, Pivovarov R, Natarajan K, Weiskopf N, Wood F, Elhadad N. Diagnosis code assignment: models and evaluation metrics. Journal of the American Medical Informatics Association. 2014;21(2):231–237.

[31] Wang Sm, Chang Yh, Kuo Lc, Lai F, Chen Yn, Yu Fy, et al. Using Deep Learning for Automatic Icd-10 Classification from FreeText Data. Eur J Biomed Inform. 2020;16(1):1–10.

[32] Lita LV, Yu S, Niculescu S, Bi J. Large scale diagnostic code classification for medical patient records. In: Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II; 2008. .

[33] Yan C, Fu X, Liu X, Zhang Y, Gao Y, Wu J, et al. A survey of automated International Classification of Diseases coding: development, challenges, and applications. Intelligent Medicine. 2022;2(3):161–173.

[34] Almeida F, Xexéo G. Word embeddings: A survey. arXiv preprint arXiv:190109069. 2019;.

[35] Kalyan KS, Sangeetha S. SECNLP: A survey of embeddings in clinical natural language processing. Journal of biomedical informatics. 2020;101:103323.

[36] Patel K, Patel D, Golakiya M, Bhattacharyya P, Birari N. Adapting pre-trained word embeddings for use in medical coding. In: BioNLP 2017; 2017. p. 302–306.

[37] dos Santos ABV, Gumiel YB, Carvalho DR. Using deep convolutional neural networks with self-taught word embeddings to perform clinical coding. Iberoamerican Journal of Applied Computing. 2018;8(1).

[38] Nath N, Lee SH, Lee I. Application of specialized word embeddings and named entity and attribute recognition to the problem of unsupervised automated clinical coding. Computers in Biology and Medicine. 2023;165:107422.

[39] Biseda B, Desai G, Lin H, Philip A. Prediction of ICD codes with clinical BERT embeddings and text augmentation with label balancing using MIMIC-III. arXiv preprint arXiv:200810492. 2020;.

[40] Yogarajan V, Gouk H, Smith T, Mayo M, Pfahringer B. Comparing high dimensional word embeddings trained on medical text to bag-of-words for predicting

medical codes. In: Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings, Part I 12. Springer; 2020. p. 97–108.

[41] Zhang S, Zhang B, Zhang F, Sang B, Yang W. Automatic ICD Coding Exploiting Discourse Structure and Reconciled Code Embeddings. In: Proceedings of the 29th International Conference on Computational Linguistics; 2022. p. 2883–2891.

[42] Steiger E, Kroll LE. Patient Embeddings From Diagnosis Codes for Health Care Prediction Tasks: Pat2Vec Machine Learning Framework. JMIR AI. 2023;2:e40755.

[43] Shi W, Wu J, Yang X, Chen N, Mien IH, Kim Jj, et al. Analyzing Code Embeddings for Coding Clinical Narratives. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021; 2021. p. 4665–4672.

[44] Bao W, Lin H, Zhang Y, Wang J, Zhang S. Medical code prediction via capsule networks and ICD knowledge. BMC Medical Informatics and Decision Making. 2021;21(2):1–12.

[45] Feucht M, Wu Z, Althammer S, Tresp V. Description-based label attention classifier for explainable ICD-9 classification. arXiv preprint arXiv:210912026. 2021;.

[46] Sezgin E, Chekeni F, Lee J, Keim S. Clinical Accuracy of Large Language Models and Google Search Responses to Postpartum Depression Questions: Cross-Sectional Study. Journal of Medical Internet Research. 2023;25:e49240.

[47] Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. Canadian Association of Radiologists Journal. 2023;p. 08465371231171125.

[48] Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elias PA, et al. Evaluating large language models on medical evidence summarization. npj Digital Medicine. 2023;6(1):158.

[49] Dong H, Falis M, Whiteley W, Alex B, Matterson J, Ji S, et al. Automated clinical coding: what, why, and where we are? NPJ digital medicine. 2022;5(1):159.

[50] Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. Jama. 2023;330(9):866–869.

[51] Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172–180.

[52] Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nature medicine. 2023;29(8):1930–1940.

[53] Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS digital health. 2023;2(2):e0000198.

[54] Karabacak M, Margetis K. Embracing Large Language Models for Medical Applications: Opportunities and Challenges. Cureus. 2023;15(5).

[55] Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obesity surgery. 2023;p. 1–7.

[56] Poon H, Naumann T, Zhang S, González Hernández J. Precision Health in the Age of Large Language Models. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2023. p. 5825–5826.

[57] Rajendran P, Zenonos A, Spear J, Pope R. A meta-embedding-based ensemble approach for ICD coding prediction. arXiv preprint arXiv:210213622. 2021;.

[58] Neelakantan A, Xu T, Puri R, Radford A, Han JM, Tworek J, et al.: Text and Code Embeddings by Contrastive Pre-Training.

[59] OpenAI.: GPT-4 Technical Report.

[60] Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Scientific data. 2018;5(1):1–13.

[61] Altman D, Machin D, Bryant T, Gardner M. Statistics with confidence: confidence intervals and statistical guidelines. John Wiley & Sons; 2013.

[62] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al.: Language Models are Few-Shot Learners.

[63] Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society series c (applied statistics). 1979;28(1):100–108.

[64] Humaira H, Rasyidah R. Determining the appropiate cluster number using Elbow method for K-Means algorithm. In: Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia; 2020. .