Validating the Application of Clinical Department-specific Artificial Intelligence-assisted Coding using TwDRGs

An-Tai Lu, Chong-Sin Liou, Chia-Hsin Lai, Bo-Tsz Shian, Ming-Ta Li, Chih-Yen Sun, Hao-Yun Kao, Hong-Jie Dai, Ming-Ju Tsai

Submitted to: JMIR Human Factors
on: April 30, 2024

Table of Contents

Validating the Application of Clinical Department-specific Artificial Intelligence-assisted Coding using TwDRGs

An-Tai Lu[1, 2] MS; Chong-Sin Liou[3] MS; Chia-Hsin Lai[3]; Bo-Tsz Shian[3]; Ming-Ta Li[1] BBA; Chih-Yen Sun[4] BS; Hao-Yun Kao[5*] PhD; Hong-Jie Dai[6, 7, 8, 9*] PhD; Ming-Ju Tsai[1, 10, 11*] PhD

[1]Department of Medical Records Kaohsiung Medical University Hospital Kaohsiung TW
[2]Department of Healthcare Administration and Medical Informatics Kaohsiung Medical University Kaohsiung TW
[3]Intelligent System Lab, College of Electrical Engineering and Computer Science Department of Electrical Engineering National Kaohsiung University of Science and Technology Kaohsiung TW
[4]Department of Internal Medicine School of Medicine, College of Medicine Kaohsiung Medical University Kaohsiung TW
[5]Department of Health Administration and Medical Informatics Kaohsiung Medical University Kaohsiung TW
[6]Department of Electrical Engineering National Kaohsiung University of Science and Technology Kaohsiung TW
[7]School of Post-Baccalaureate Medicine Kaohsiung Medical University Kaohsiung TW
[8]National Institute of Cancer Research National Health Research Institutes Tainan TW
[9]Center for Big Data Research Kaohsiung Medical University Kaohsiung TW
[10]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine Kaohsiung Medical University Hospital Kaohsiung TW
[11]Department of Medical Technology Kaohsiung Medical University Hospita Kaohsiung TW
[*]these authors contributed equally

Corresponding Author:
Hao-Yun Kao PhD
Department of Health Administration and Medical Informatics
Kaohsiung Medical University
No.100, Shih-Chuan 1st Road, Sanmin Dist.
Kaohsiung
TW

Abstract

Background: The accuracy of ICD-10-CM/PCS (International Classification of Diseases, 10th Revision Clinical Modification/Procedure Coding System) coding is crucial for generating correct Taiwan Diagnosis-Related Groups (TwDRGs), as coding errors can lead to financial losses for hospitals.

Objective: The study aimed to determine the consistency between the artificial intelligence (AI)-assisted coding module and manual coding, as well as identifying clinical specialties suitable for implementing the developed AI-assisted coding module.

Methods: This study validates the AI-assisted coding module from the perspective of healthcare professionals. The research period commenced in February 2023. The study subjects excluded cases outside of TwDRGs, those with incomplete medical records, and cases with TwDRGs disposals ICD-10-PCS. Data collection was conducted through retrospective medical record review. The AI-assisted module was constructed using a hierarchical attention network (HAN). The verification of the TwDRGs results from the AI-assisted coding model focused on the major diagnostic category (MDC). Statistical computations were conducted using statistical package for the social sciences (SPSS) software, while research variables consisted of categorical variables represented by MDC, and continuous variables represented by the RW of TwDRGs.

Results: A total of 2,632 discharge records meeting the research criteria were collected from 0February to April 2023. In terms of inferential statistics, Kappa statistics were employed for MDC analysis. The infectious diseases, parasitic diseases and respiratory system had Kappa values exceeding 0.8. Clinical inpatient specialties were statistically analyzed using the Wilcoxon Signed Rank Test. There was no difference in coding results between 23 clinical departments such as Division of Cardiology, Division of Nephrology, and Department of Urology classification personnel.

Conclusions: For human coders, with the assistance of the ICD-10-CM/PCS AI-assisted coding system, work time is reduced; additionally, strengthening knowledge in clinical documentation improvement (CDI) enables human coders to maximize their role. This positions them to become CDI experts1, preparing them for further career development. Future research will apply the same methodology to validate the ICD10PCS AI-assisted coding module.

(JMIR Preprints 30/04/2024:59961)
DOI: https://doi.org/10.2196/preprints.59961

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

An-Tai Lu, MS[1,2], Chong-Sin Liou, MS[3], Chia-Hsin Lai[3], Bo-Tsz Shian[3], Ming-Ta Li ,BBA[2] Chih-Yen Sun,BS[4], Hao-Yun Kao,PhD[1]*, Hong-Jie Dai,PhD [3,5,6,7]*, Ming-Ju Tsai,MD,PhD[2,8,9]*

[1]Department of Healthcare Administration and Medical Informatics, College of Health Sciences, Kaohsiung Medical University, Kaohsiung , Taiwan

[2]Department of Medical Records, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[3]Intelligent System Lab, College of Electrical Engineering and Computer Science, Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[4]Department of Internal Medicine, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[5]School of Post-Baccalaureate Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[6]National Institute of Cancer Research, National Health Research Institutes, Tainan, Taiwan

[7]Center for Big Data Research, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

[8]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

[9]Department of Medical Technology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

* Corresponding author: haoyun@kmu.edu.tw, tel: +886-7-3121101 ext. 2648#23; hjdai@nkust.edu.tw, tel: +886-7-381-4526 ext. 15510; mjt@kmu.edu.tw, tel: +886-7-3121101 ext. 5651;

Validating the Application of Clinical Department-specific Artificial Intelligence-assisted Coding using TwDRGs

Abstract

Background:

The accuracy of ICD-10-CM/PCS (International Classification of Diseases, 10th Revision Clinical Modification/Procedure Coding System) coding is crucial for generating correct Taiwan Diagnosis-Related Groups (TwDRGs), as coding errors can lead to financial losses for hospitals.

Objective:

The study aimed to determine the consistency between the artificial intelligence (AI)-assisted coding module and manual coding, as well as identifying clinical specialties suitable for implementing the developed AI-assisted coding module.

Methods:

This study validates the AI-assisted coding module from the perspective of healthcare professionals. The research period commenced in February 2023. The study subjects excluded cases outside of TwDRGs, those with incomplete medical records, and cases with TwDRGs disposals ICD-10-PCS. Data collection was conducted through retrospective medical record review. The AI-assisted module was constructed using a hierarchical attention network (HAN). The verification of the TwDRGs results from the AI-assisted coding model focused on the major diagnostic category (MDC). Statistical computations were conducted using statistical package for the social sciences (SPSS) software, while research variables consisted of categorical variables represented by MDC, and continuous variables represented by the RW of TwDRGs.

Results:

A total of 2,632 discharge records meeting the research criteria were collected from 0February to April 2023. In terms of inferential statistics, Kappa statistics were employed for MDC analysis. The infectious diseases, parasitic diseases and respiratory system had Kappa values exceeding 0.8. Clinical inpatient specialties were statistically analyzed using the Wilcoxon Signed Rank Test. There was no difference in coding results between 23 clinical departments such as Division of Cardiology, Division of Nephrology, and Department of Urology classification personnel.

Conclusions:

For human coders, with the assistance of the ICD-10-CM/PCS AI-assisted coding system, work time is reduced; additionally, strengthening knowledge in clinical documentation improvement (CDI) enables human coders to maximize their role. This positions them to become CDI experts[1], preparing them for further career development. Future research will apply the same methodology to validate the ICD10PCS AI-assisted coding module.

Keywords: ICD-10-CM, TwDRGs, artificial intelligence-assisted coding

Introduction

The International Classification of Diseases (ICD) system was established by the World Health Organization (WHO) for the purpose of tracking diseases globally. Over the past several decades, the WHO has made significant changes to both content and structure. It accompanies new scientific understanding of disease and new structures for organizing ICD codes to accommodate enhanced usage and extensibility[2]. The WHO introduced the ICD in 1948, which is a universal language used to categorize diseases or causes of death. The use of it is attributed to healthcare-related units in 194 countries, and is generated by professional coders based on discharge records, with countries adjusting ICD to their circumstances. In 2016, Taiwan adopted the international trend of switching from ICD-9-CM (Clinical Modification) to ICD-10-CM/PCS (Procedure Coding System) for coding hospital patient diagnoses, procedures, analysis and reimbursement. The National Health Insurance Administration (NHIA) under the Ministry of Health and Welfare has adopted the 2014 edition of ICD-10-CM/PCS, with approximately 71,900 diagnosis codes and 78,500 procedure codes.

The use of it involves coding and classifying morbidity data from health records, reimbursement claims, and administrative databases. Improving health care quality, monitoring public health and conducting research are all benefits of ICD-10-CM in Taiwan, involving converting the physician's discharge diagnosis into ICD-10-CM/PCS codes by following the primary diagnosis selection principle announced by the NHIA. The diagnosis-related group (DRG) provides information such as health insurance reimbursement, relative weight, presence of comorbidities, and complications for the current hospitalization.

The accuracy of ICD-10-CM/PCS coding is crucial for generating accurate TwDRGs, as coding errors can lead to financial losses for hospitals [3,4]. According to the coding principles set forth by the NHIA and the Taiwan Society of Medical Records Management, coding is based on the inpatient and emergency room records of patients. In the past, this task was undertaken by trained and certified clinical coding professionals (referred to as "coding professionals" hereafter), but with the rapid advances of medical technology, the rules of disease classification also evolve, indicating professional coding professionals must regularly accumulate relevant training hours to update their disease classification skills[5].

In recent years, artificial intelligence (AI) and natural language processing (NLP) have shown great potential in the field of automatic clinical coding. In 2021, the disease coding scales in the United States were worth approximately 18 billion USD. Several technology companies in the U.S. have developed AI-assisted coding systems, and scholars believe that interdisciplinary collaboration and feedback from clinical coding professionals are essential to further refine the modules[6 7]. Research on AI-assisted coding consistently concludes that it improves quality and reduces error rates while saving costs[8,9]. ICD-10-CM/PCS AI-assisted coding can be considered as a text classification task within the field of machine learning (ML). In recent years, studies in the ML text classification field have predominantly proposed using deep learning-based neural networks[10]. The development and validation process of the AI-assisted coding model requires the involvement and feedback of clinical coders to enhance accuracy and correctness, aligning with user needs[1].

In Taiwan, several hospitals have also ventured into the development of AI-assisted coding for disease classification; however, due to variations in physicians' documentation of medical records across different hospitals, the AI -assisted coding systems developed are not universally applicable[11], necessitating the development and validation of customized AI-assisted coding systems.

Medical coding personnel are required to review the discharge records meticulously, and then translate the discharge diagnoses and procedures (interventions) recorded in the medical records into ICD-10 codes. In the past, the most significant factor contributing to coding errors was handwritten medical records by physicians, which were difficult to decipher or included abbreviations, leading to mistakes[12]. In recent years, the majority of medical centers in Taiwan have adopted electronic health records, resulting in a significant reduction in coding errors caused by such handwritten records. Clinical coding personnel also encounter various pressures, including the need to accomplish all inpatient coding tasks within specified deadlines, optimize TwDRGs assignment coding, enhance and maintain coding reliability and validity, and engage in discussions with clinical physicians regarding the content of medical record writing.

In recent times, the global trend in AI coding has been on the rise[11,13-16]. In this study, we have developed an exclusive ICD-10-CM AI-assisted coding module. Coding professionals were involved in the research and offered suggestions to improve the efficiency of coding operations. Consequently, this study focuses primarily on the following two research aims:

1. To verify the consistency between the AI-assistant coding module and the coding professional in encoding based on the MDC results in TwDRGs.
2. To identify the clinical departments within the medical center that are beneficial for employing the developed AI-assisted coding module.

Methods

Data Description

This study verifies the AI-assisted coding from the perspective of coding professionals. Since the AI-assisted coding system was introduced in the medical center in February of the year 2023, the study period commenced from February 2023. The subjects of this study were selected based on the exclusion criteria of non-Tw-DRGs cases, cases with procedures (ICD-10-PCS) in TwDRGs, and cases with incomplete medical records. According to the study conditions, there were approximately 700 to 1,000 cases per month. The coding by both the AI-assisted coding module and coding professionals were based on the electronic discharge summaries of a certain medical center each month.

Research Design

After each data entry was encoded by the AI-assisted coding module and verified by a coding professional, it was transmitted to a certain university's database. The results of both the AI-assisted coding and coding professional were compared using an Excel file. Following the linkage to the National Health Insurance Administration's DRG calculation software, separate datasets for TwDRGs were obtained for both the AI-assisted coding and coding personnel, with the consistency of the primary diagnosis coding between these two being examined. In cases of discrepancies, the medical records were scrutinized again by the coding professional to determine if the AI-assisted coding results met the criteria for primary coding as per the coding professional; accordingly, the consistency results of TwDRGs data for both the AI-assisted coding and coding professional were adjusted.

AI -Assisted Coding Construction Process

The AI-assisted ICD-10-CM coding system was developed by CSL, CHL, and BTS based on approximately 110,000 discharge summaries collected from April 1, 2019, to December 31, 2020 in a medical center. The de-identified summary data were processed by segmenting sentences and filtering out meaningless delimiters and prefix symbols (e.g., # or "") by employing a clinical NLP tool[17] . The data was categorized into 21 groups based on the first three codes of ICD-10-CM, and models were built using Bidirectional Encoder Representations from Transformers (BERT)[18] and hierarchical attention network (HAN)[19]. The results favored HAN, leading to the decision to adopt the HAN module. The precision, recall and F1-scores of the developed HAN model are 0.55, 0.82, and 0.66 respectively. For top-50 most frequent codes, the F-score of the developed HAN model is 0.818.

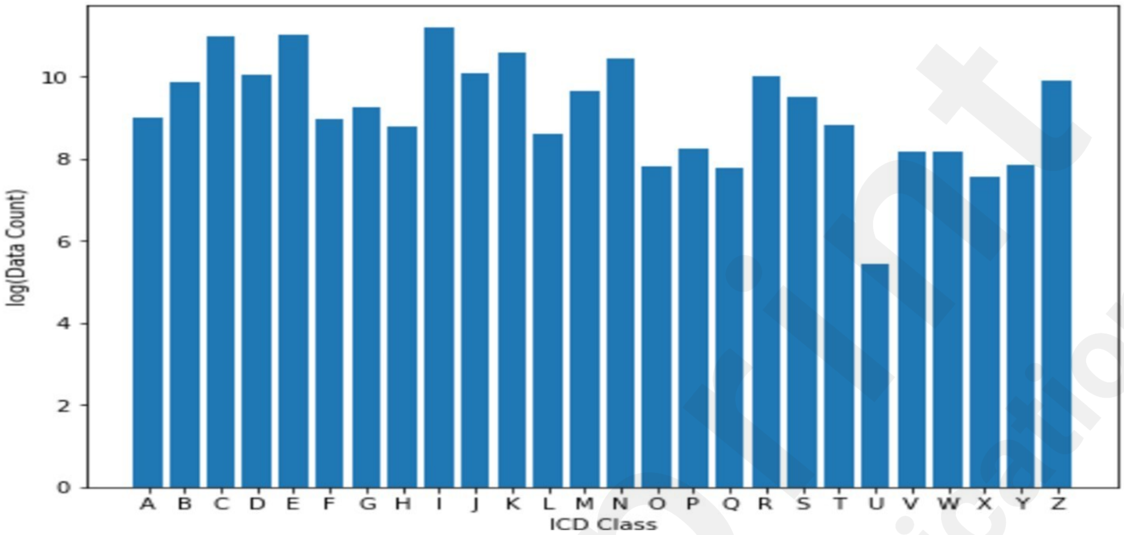

Figure 2: The volume of the collected dataset (Source: Liu, C., 2023). The y-axis represents the logarithmically scaled number of acquired data entries.

The distribution of the ICD10 codes observed in the collected training dataset is shown in Figure 2. The first digit of the ICD-10-CM code consists of English letters, so the alphabetical characters on the horizontal axis of the log data represent the first digit of the ICD-10-CM code, indicating diseases pertaining to different systems. According to Figure 2, data starting with codes C, E, and I in ICD-10-CM have the highest volume, with C representing neoplastic, E for endocrinal, nutritional and metabolic diseases, and I for diseases of the circulatory system respectively. These are the body systems with the highest learning volumes for the AI-assisted coding module.

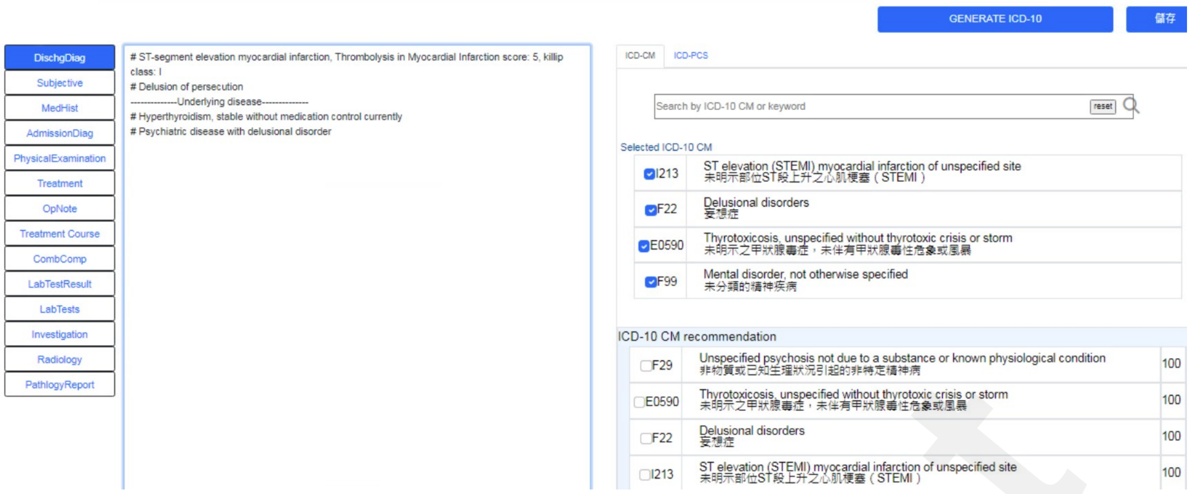

Figure 3 User Interface Illustration

Aside from module modeling, another time-consuming task was the design of the user interface for coding professional, as it needed to present discharge summaries, laboratory data, imaging reports, as well as the ICD-10-CM codes predicted by the AI-assisted coding module. Coding professionals were actively involved in providing feedback during the interface design process. Figure 3 provides an illustration of the designed user interface, which provides suggestions automatic ICD-10-CM Coding and fields for coding professional to input the final codes. The developed AI-coding system was integrated into a medical center's hospital information systems in November 2022, and operated in February 2023.

Results

The study involves an analysis incorporating descriptive statistics for exploration, as well as inferential statistics for investigating MDC and relative weight (RW). Statistical computations were conducted using statistical package for the social sciences (SPSS) version 19 software, while research variables consisted of categorical variables represented by MDC, and continuous variables represented by the RW of TwDRGs.

Descriptive Statistics Section

In the period from February to April of the year 2023, a total of 15,756 discharges were recorded. Excluding cases with interventions, non-TwDRGs cases, and cases with incomplete medical records, there were a total of 2,632 cases. The primary diagnosis was the key factor in determining the main disease category, while secondary diagnoses only affected the distribution of TwDRGs within the same primary disease category. According to disease classification rules, the primary diagnosis was based on the reason for the patient's admission, but only one disease could be selected as the primary diagnosis. If multiple diseases were treated during admission, selecting any one of them as the primary diagnosis was not considered an error. Therefore, the coding professional, ATL, manually examined the discharged cases' notes to categorize the output of AI-assisted system into one of the following category. The results are shown in Table 1.

Table 1: Frequency Distribution and Percentage Analysis of Primary Diagnoses

	February N (%)	March N (%)	April N (%)
No Primary Diagnosis	181(24.2)	277(28.0)	285(31.9)
Incorrect Secondary Diagnosis with a Primary Diagnosis	462(61.8)	477(48.1)	369(41.3)
All Correct	79(10.6)	181(18.3)	177(19.8)
All Incorrect	26(3.5)	56(5.7)	62(7.0)
Total	748(100)	991(100)	893(100)

Operational definitions are as follows:

• No Primary Diagnosis: In comparison to the coding professional, a single hospitalization's predicted diagnosis codes does not include a primary diagnosis.

• Incorrect Secondary Diagnosis with a Primary Diagnosis: In comparison to the coding professional, a single hospitalization's predicted diagnosis codes includes a primary diagnosis, but there is at least one error in the secondary diagnoses.

• All Correct: All predicted diagnosis codes for a single hospitalization perfectly align with those given by the coding professional.

• All Incorrect: In comparison to the disease classification personnel, none of the predicted diagnosis codes in a single hospitalization are the same.

In Figure 4, we analyzed the agreement of MDC between the AI-assisted coding module and the coding professional through the heatmap analysis. The vertical and horizontal axes in Figure 3 represent MDCs coded by the AI-Assisted coding module, and MDCs coded by coding professionals, respectively. The intensity of color in the figure indicates a higher number of agreed MDCs between the AI-assisted coding module and professionals. As shown in Figure 3, MDC 1 (Diseases and Disorders of the Nervous System), MDC 4 (Diseases and Disorders of the Respiratory System) and MDC 18 (Infectious and Parasitic Diseases and Disorders) appear having the highest agreements.

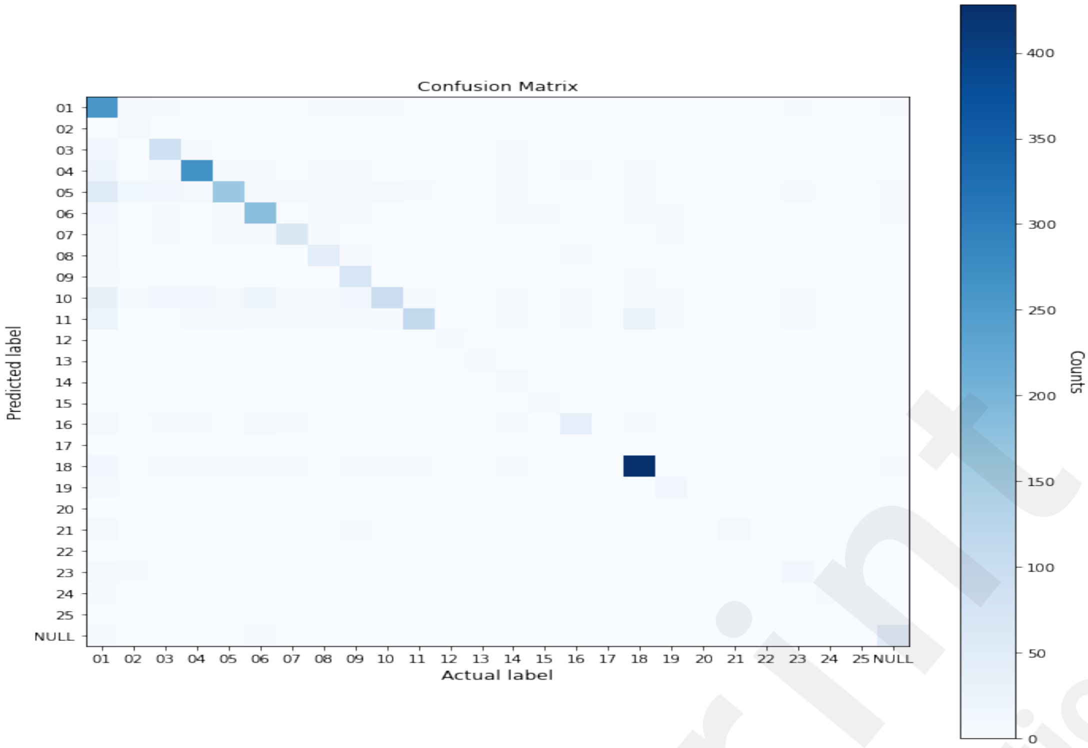

Figure 4: MDC heatmap analysis between AI-coding module and professionals.

Kappa Coefficient Test:

Furthermore, we assess the MDC agreement between the AI coding module and coding professionals in terms of the Kappa coefficient test. The Kappa values are broadly categorized into five groups to represent different levels of agreement:

- 0.0 ~ 0.20: Extremely low agreement
- 0.21 ~ 0.40: Fair agreement
- 0.41 ~ 0.60: Moderate agreement
- 0.61 ~ 0.80: High agreement
- 0.81 ~ 1: Almost perfect agreement

When analyzing the cumulative data for February to April 2023, as presented in Table 2, the MDCs with the highest consistency were MDC 4 (Diseases and Disorders of the Respiratory System) and MDC 18 (Infectious and Parasitic Diseases and Disorders), followed by MDC 1 (Diseases and Disorders of the Nervous System), MDC 3 (Diseases and Disorders of the Ear, Nose, Mouth And Throat), MDC 6 (Diseases and Disorders of the Digestive System), MDC 7 (Diseases and Disorders of the Hepatobiliary System And Pancreas), MDC 9 (Diseases and Disorders of the Skin, Subcutaneous Tissue And Breast), MDC 11 (Diseases and Disorders of the Kidney And Urinary Tract), MDC 13 (Diseases and Disorders of the Female Reproductive System), MDC 15 (Newborn And Other Neonates (Perinatal Period)), MDC 16(Diseases and

Disorders of the Blood and Blood Forming Organs and Immunological Disorders).

Table 2□Kappa Tests for Aggregation of MDC in the Total Counts for February to April 2023

Major Diagnostic Category		AI-assisted case coding count (%)	The number of cases coded by human coders (%)	Kappa value	P value
1	Diseases and Disorders of the Nervous System	280(10.6)	509(19.3)	0.67□	P<0.05
2	Diseases and Disorders of the Eye	9(0.3)	38(1.4)	0.3	P<0.05
3	Diseases and Disorders of the Ear, Nose, Mouth and Throat	113(4.3)	132(5.0)	0.689□	P<0.05
4	Diseases and Disorders of the Respiratory System	309(11.7)	302(11.5)	**0.845□□**	**P<0.05**
5	Diseases and Disorders of the Circulatory System	310(11.8)	184(7.0)	0.607	P<0.05
6	Diseases and Disorders of the Digestive System	217(8.2)	229(8.7)	0.775□	P<0.05
7	Diseases and Disorders of the Hepatobiliary System and Pancreas	90(3.4)	84(3.2)	0.710□	P<0.05
8	Diseases and Disorders of the Musculoskeletal System and Connective Tissue	66(2.5)	87(3.3)	0.576	P<0.05
9	Diseases and Disorders of the Skin, Subcutaneous Tissue and Breast	83(3.2)	116(4.4)	0.692□	P<0.05
1	Diseases and Disorders	237(9.0)	132(5.0)	0.505	P<0.05

0	of the Endocrine, Nutritional and Metabolic Systems				
11	Diseases and Disorders of the Kidney and Urinary Tract	205(7.8)	120(4.6)	0.648□	P<0.05
12	Diseases and Disorders of the Male Reproductive System	7(0.3)	3(0.1)	0.362	P<0.05
13	Diseases and Disorders of the Female Reproductive System	8(0.3)	8(0.3)	0.749□	P<0.05
14	Pregnancy, Childbirth and Puerperium	11(0.4)	41(1.6)	0.419	P<0.05
15	Newborn And Other Neonates (Perinatal Period)	9(0.3)	15(0.6)	0.635□	P<0.05
16	Diseases and Disorders of the Blood and Blood Forming Organs and Immunological Disorders	64(2.4)	57(2.2)	0.624□	P<0.05
17	Myeloproliferative Diseases and Disorders (Poorly Differentiated Neoplasms)	3(0.1)	2(0.1)	0.399	P<0.05
18	Infectious and Parasitic Diseases and Disorders	465(17.7)	505(19.2)	**0.870□□**	**P<0.05**
19	Mental Diseases and Disorders	23(0.9)			
20	Alcohol or drug abuse or induced organic mental disorder	2(0.1)			
21	Injuries, Poison, and Toxic Effect of Drugs	24(0.9)	21(0.8)	0.404	P<0.05
2	Burns		1(0.0)	-	

2					
23	Factors Influencing Health Status and Other Contacts with Health Services	24(0.9)	42(1.6)	0.366	P<0.05
24	Multiple Significant Trauma	6(0.2)	4(0.2)	0.599	P<0.05
25	HIV infection	2(0.1)			
	None	65(2.5)			
	Total	2632(100)	2632(100)	0.592	

• 0.61 ~ 0.80: High agreement

• 0.81 ~ 1: Almost perfect agreement

Inferential Statistical Analysis: Wilcoxon Signed Rank Test

The Kappa coefficient test belongs to a broad-scale analysis of MDC. However under the same MDC, it is possible to further classify the data into numerous TwDRGs, with each having its own code and RW. Even with the same MDC, this might result in different TwDRGs. Furthermore, some diseases can be treated across departments. Therefore, for the statistical analysis of RW, we first conduct a normality analysis of the RW obtained from both AI-assisted coding and coding professionals. The statistical results based on the Kolmogorov-Smirnov analysis yields a significance level (P-value) less than 0.05, indicating a non-normal distribution. Given that the research sample consists of paired data, the non-parametric Wilcoxon signed rank test (WSRT) is employed to analyze whether there were differences in RW between AI-assisted coding and coder-assigned coding, the null hypothesis assumes that there is no difference in RW between AI-assisted coding and coder-assigned coding.

WSRT, with clinical departments as the unit of analysis, identified differences in RW in the following departments: Division of Endocrinology and Metabolism, Division of Hematology and Oncology, Division of General Internal Medicine, Division of Geriatrics and Gerontology, Division of Trauma, Division of Neurosurgery, Division of Cardiovascular Surgery, Division of General and Digestive Surgery, Division of Pediatric Neurology, Department of Otorhinolaryngology, Department of Neurology, and Department of Rehabilitation Medicine, totaling 12 departments. As shown in Table 3, the overall statistical result with a p-value less than 0.05 indicates that there are still differences between AI-assisted coding and coder-assigned coding at present.

Table 3：Wilcoxon Signed-Rank Test results across various clinical departments

Clinical	Count (Percentage)	P Value
Division of Gastroenterology	61(2.3)	P=0.115
Division of Hepatobiliary and Pancreatic Medicine	81(3.1)	P=0.073
Division of Cardiology	177(6.7)	P=0.390
Division of Chest Medicine	181(6.9)	P=0.503
Division of nephrology	64(2.4)	P=0.906
Division of Endocrinology and Metabolism	**36(1.4)**	**P=0.030**＊
Division of Hematology and Oncology	**42(1.6)**	**P=0.006**＊
Division of Rheumatology, Immunology and Allergology	36(1.4)	P=0.332
Division of Infectious Diseases	78(3.0)	P=0.394
Division of General Internal Medicine	**205(7.8)**	**P=0.005**＊
Division of Geriatrics and Gerontology	**54(2.1)**	**P=0.003**＊
Division of Trauma	**16(0.6)**	**P=0.038**＊
Division of Neurosurgery	**151(5.7)**	**P<0.000**＊
Division of Cardiovascular Surgery	**24(0.9)**	**P=0.002**＊
Division of Chest Surgery	14(0.5)	P=0.397
Division of Pediatric Surgery	9(0.3)	P=0.285
Division of Plastic Surgery	9(0.3)	P=0.893
Division of Colorectal Surgery	46(1.7)	P=0.156
Division of Breast Oncology and Surgery	16(0.6)	P=0.05
Division of General and Digestive Surgery	**56(2.1)**	**P=0.009**＊
Department of Gynecology Obstetrics	60(2.3)	P=0.722
Division of Pediatric Hematology and Oncology	42(1.6)	P=0.443
Division of Pediatric Cardiology and Pulmonology	86(3.3)	P=0.244
Division of Pediatric	**92(3.5)**	**P=0.008**＊

Neurology		
Division of Neonatology	14(0.5)	P=0.686
Division of General Pediatrics	299(11.4)	P=0.558
Division of Pediatric Allergy Immunology	8(0.3)	P=0.317
Department of Otorhinolaryngology	**53(2.0)**	**P=0.020**
Ophthalmology Department	12(0.5)	P=0.441
Department of Orthopaedics	13(0.5)	P=0.161
Department of Urology	46(1.7)	P=0.950
Department of Dermatology	87(3.3)	P=0.175
Department of Neurology	**366(13.9)**	**P<0.000**
Division of Family Medicine	49(1.9)	P=0.173
Department of Rehabilitation Medicine	**47(1.8)**	**P<0.000**
Department of Psychiatry	1(0.0)	**-**
Division of Oral Maxillofacial Surgery	1(0.0)	**-**
Total	2632	**P<0.000**

P value< 0.05

Discussion

Principal Results

For clinical coders, it is evident from the MDC that AI-assisted coding can serve as a reference for disease systems. However, hospital administrators may require detailed statistical results from clinical departments to make appropriate judgments. In the individual clinical department analysis based on WSRT, the Division of General Internal Medicine, Department of Neurology and Division of Neurosurgery had the highest number of cases studied, but the statistical results were inconsistent with coder-assigned coding, although in the Kappa coefficient test, the statistical results for the nervous system were highly consistent. This is because patients admitted to the Department of Neurology and Neurosurgery do not exclusively have neurological conditions; they might also have circulatory system disorders such as strokes or intracranial hemorrhages. This accounts for the discrepancies observed in both the WSRT and the Kappa coefficient test.

In the circulatory system, the statistical results for the Division of Cardiology and Division of Cardiovascular Surgery in the WSRT were also markedly different. Upon closer examination of the data from the exploratory study, it was discovered that in the Division of Cardiovascular

Surgery, half of the cases assisted by AI-coding modules did not have the main diagnosis coded, which could be attributed to differences in how physicians document medical records.

Furthermore, in MDC 14 (Pregnancy, Childbirth, and Puerperium) and MDC 21 (Injuries, Poisonings, and Toxic Effects of Drugs), there are specific coding rules. Clinical coders need to synthesize the entire medical record information and apply the coding rules, which could result in diagnoses different from those presented in the discharge summary.

Limitations

The AI-coding assistance module was trained on inpatient data from April 2019 to December 2020. Advancements in medical care might lead to variations in the diseases of admitted patients. Taken together, these indicate situations where the AI-coding assistance module might not capture the main diagnosis, as observed in the Dermatology Department.

Conclusions

This study had two main objectives. The first was to investigate the consistency between the AI-assisted coding module and coding professionals, and the second was to identify the departments suitable for utilizing the AI-assisted coding module. The research results indicated that the highest consistency in MDC was observed in diseases of the respiratory system and infectious and parasitic diseases. In the analysis of various inpatient specialties, departments such as the Division of Cardiology, Division of Nephrology, and Department of Urology showed no significant difference from coder-assigned coding results; accordingly, consideration could be given to integrating the AI-assisted coding module into the hospital information system, allowing physicians to reference TwDRGs assignments for hospitalized patients, thus effectively controlling medical expenses.

However, upon analyzing the entire hospital department, discrepancies were observed in alignment with disease categorizations and personnel coding, so the research team is actively working on continuous improvements; nevertheless, the AI-assisted coding indeed served as a valuable reference, reducing human errors, as during the research period, it was found that the error rate detected by human coders (number of coding errors by human coders / total cases) was 1.31%. Given the regular updates to the toolbook by the Department of Health and the revisions in coding rules, the use of the coding assistance module undoubtedly proves to be a powerful tool.

The development of AI-assisted coding for ICD-10-CM/PCS is just the beginning for intelligent healthcare in disease classification. Many operational aspects of hospitals are closely related to ICD-10=CM/PCS, including inpatient coding monitoring, discharge preparation services, and infectious disease surveillance, among others. For hospital administrators, the ultimate goal of AI-assisted coding is to achieve optimal operational revenue.

For human coders, with the assistance of the ICD-10-CM/PCS AI coding system, work time is reduced; additionally, strengthening knowledge in CDI enables human coders to maximize their role, positioning them to become CDI experts [1] and preparing them for further career development.

Acknowledgements

This study was supported by grants from Kaohsiung Medical University Hospital (SI11212)

Conflicts of Interest

none declared..

Abbreviations

AI: artificial intelligence

ICD10CM/PCS: International Classification of Diseases, 10th Revision Clinical Modification/Procedure Coding System

TwDRGs: Taiwan Diagnosis Related Group

MDC: Major Diagnostic Category

WHO: World Health Organization

NHIA: National Health Insurance Administration

RW: Relative Weight

ML: Machine Learning

DL: Deep Learning

BERT: bidirectional encoder representations from transformers

CDI: Clinical Documentation Improvement

HAN: Hierarchical Attention Networks

UI: User Interface

Uncategorized References

1. Campbell S, Giadresco K. Computer-assisted clinical coding: A narrative review of the

literature on its benefits, limitations, implementation and impact on clinical coding professionals. *Health Inf Manag*. Jan 2020;49(1):5-18. doi:10.1177/1833358319851305
2. Hirsch JA, Nicola G, McGinty G, et al. ICD-10: History and Context. *AJNR Am J Neuroradiol*. Apr 2016;37(4):596-9. doi:10.3174/ajnr.A4696
3. Ayub S, Scali ST, Richter J, et al. Financial implications of coding inaccuracies in patients undergoing elective endovascular abdominal aortic aneurysm repair. *J Vasc Surg*. Jan 2019;69(1):210-218. doi:10.1016/j.jvs.2018.04.027
4. Toner E, Khaled A, Ramesh A, Qureshi MK, Al Suyyagh K, Dunkow P. Financial Impact of Inaccurate Coding Plus Cost-Effectiveness Analysis for Surgically Managed Patients With Periprosthetic Fractures. *Cureus*. Feb 1 2021;13(2):e13060. doi:10.7759/cureus.13060
5. Otero Varela L, Doktorchik C, Wiebe N, et al. International Classification of Diseases clinical coding training: An international survey. *Health Inf Manag*. Jul 15 2022:18333583221106509. doi:10.1177/18333583221106509
6. *U.S. Medical Coding Market Size, Share & Trends Analysis Report By Classification System (ICD, HCPCS, CPT), By Component, By Medical Specialty, And Segment Forecasts, 2023 - 2030*. https://www.grandviewresearch.com/industry-analysis/us-medical-coding-market
7. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. *NPJ Digit Med*. Feb 3 2023;6(1):16. doi:10.1038/s41746-023-00768-0
8. Tong S, Jia N, Luo X, Fang Z. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*. 2021;42(9):1600-1631. doi:10.1002/smj.3322
9. Vrontis D, Christofi M, Pereira V, Tarba S, Makrides A, Trichina E. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *The International Journal of Human Resource Management*. 2021;33(6):1237-1266. doi:10.1080/09585192.2020.1871398
10. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc*. Mar 1 2020;27(3):457-470. doi:10.1093/jamia/ocz200
11. Chen PF, Wang SM, Liao WC, et al. Automatic ICD-10 Coding and Training System: Deep Neural Network Based on Supervised Learning. *JMIR Med Inform*. Aug 31 2021;9(8):e23230. doi:10.2196/23230
12. Asadi a. Factors Affecting Clinical Coding Errors. *Shiraz E-Medical Journal*. 2022;23(9)doi:10.5812/semj-122161
13. Islam MM, Li GH, Poly TN, Li YJ. DeepDRG: Performance of Artificial Intelligence Model for Real-Time Prediction of Diagnosis-Related Groups. *Healthcare (Basel)*. Nov 25 2021;9(12)doi:10.3390/healthcare9121632
14. Shuai Z, Xiaolin D, Jing Y, et al. Comparison of different feature extraction methods for applicable automated ICD coding. *BMC Med Inform Decis Mak*. Jan 12 2022;22(1):11. doi:10.1186/s12911-022-01753-5
15. Subotin M, Davis AR. A method for modeling co-occurrence propensity of clinical codes with application to ICD-10-PCS auto-coding. *J Am Med Inform Assoc*. Sep 2016;23(5):866-71. doi:10.1093/jamia/ocv201
16. Diao X, Huo Y, Zhao S, et al. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. *Int J Med Inform*. Sep 2021;153:104543. doi:10.1016/j.ijmedinf.2021.104543
17. Chang NW, Dai HJ, Jonnagaddala J, Chen CW, Tsai RT, Hsu WL. A context-aware approach for progression tracking of medical concepts in electronic medical records. *J Biomed Inform*. Dec 2015;58 Suppl(Suppl):S150-s157. doi:10.1016/j.jbi.2015.09.013

18. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019:
19. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. *Hierarchical Attention Networks for Document Classification*. 2016:1480-1489.