IJPPR
INTERNATIONAL JOURNAL OF PHARMACY & PHARMACEUTICAL RESEARCH
An official Publication of Human Journals
ISSN 2349-7203
HUMAN

Human Journals
Review Article
February 2021 Vol.:20, Issue:3

Medical Coding and Its Uses for Diagnosis Prediction in Health Care System Management

IJPPR
INTERNATIONAL JOURNAL OF PHARMACY & PHARMACEUTICAL RESEARCH
An official Publication of Human Journals
ISSN 2349-7203
HUMAN

Kirteebala Pawar*, Rishikesh. R. Sharma[1], Shivam. D. Sharma[2], Sagar. N .Sherkar[3], Dipak. D. Shivgan[4], Dhawal. R. Shinde[5], Smita Takarkhede

Ideal College Of Pharmacy And Research, Bhal, Kalyan -421306, Dist – Thane Maharashtra, India
Affiliation To University Of Mumbai

Submitted: 03 January 2021
Revised: 23 January 2021
Accepted: 12 February 2021

Keywords: Medical coding, International Classification of Diseases (ICD), Current Procedure Terminology (CPT), Artificial intelligence (AI), Electronic Health Records

ABSTRACT

Medical coding is the process of converting information on healthcare diagnoses, medical services, or procedures into a numeric or alphanumeric format contained inpatient medical records.[1] Assigning a code for a specific diagnosis or procedure provides a way of standardizing the recording of clinical information. This recorded documentation includes such as images, medical text - physician notes, physiological signals, research laboratory, etc. However, the processing of medical record is difficult because they are structurally free, varied in style, and has subjective factors. Assigning metadata codes from the International Classification of Diseases (ICD) presents a standardized way of indicating diagnoses and procedures, so it becomes a mandatory process for understanding medical records to make better clinical and economic decisions.[2]

INTRODUCTION

1.1 Background on Medical Coding

Widespread usage of the healthcare system, massive data are continuously generated by the patients' hospitalization, which keeps each patient encounter in Electronic Health Records (EHRs), Hospital Information System (HIS), Laboratory Information System (LIS), and many other cyber-physical systems.[2]

Electronic Healthcare Records (EHR) helps in a computerized patient record, makes it possible to directly predict patients' future health status, which is done by analyzing their historical visit records.[3] [4]

Hospitals and in general healthcare providers rely on medical coding to record medical services.

Thus, healthcare providers are largely required to use medical codes to classify the present conditions/diseases and procedures performed on basically all patients. The automatic processing of this information, mainly providing support for quality care, also aims at ensuring optimal management of hospital resources.[5]

Use of Medical coding in healthcare, support to improve both the quality of healthcare services and cost control.[6] Using intelligent techniques for analyzing rich textual medical data has become mainstream in healthcare.

The current standard coding system is the International Classification of Diseases (ICD) which is a repository is maintained by the World Health Organization (WHO) to provide a hierarchy of diagnostic codes of diseases, disorders, injuries, signs, symptoms, etc.[2]

Let's start with a simple question about medical coding: Why do we code medical reports?

According to the Centers for Disease Control (CDC), there were over 1.2 billion patient visits in the past year. That's a stat that includes visits to physician offices, hospital outpatient facilities, and emergency rooms. If there were just five pieces of coded information per visit, which is an almost unrealistically low estimate, that'd be 6 billion individual pieces of information that need to be transferred every year. In a system loaded with data, medical coding allows for the efficient transfer of huge amounts of information.[7]–[9]Coding also

allows for uniform documentation between medical facilities. This coding is universally accepted and followed by the nation. Having uniform data allows for efficient research and analysis, which government and health agencies use to track health trends much more efficiently. [8], [9]

Finally, coding allows administrations to look at the prevalence and effectiveness of treatment in their facility. This is especially important to large medical facilities like hospitals. Like government agencies tracking, say, the incidence of a certain disease, medical facilities can track the efficiency of their practice by analyzing.[10]

As you can see, medical coding simplifies the business of health considerably. Now that we understand the importance of this practice, let's take a look at the three types of code that you'll have to become familiar with as a medical coder.

There are three sets of code that you will use daily as a medical coder.

1. **International Classification of Diseases or (ICD) codes.**
2. **Current Procedure Terminology, or CPT**
3. **Healthcare Common Procedure Coding System (HCPCS)**

1. International Classification of Diseases

These are diagnostic codes that create a uniform vocabulary for describing the causes of injury, illness, and death. This code set was established by the World Health Organization in the late 1940s. It's been updated several times in the 60-plus years since its inception. The number following "ICD" represents which revision of the code is in use.

ICD codes are used to represent a doctor's diagnosis and the patient's condition. In the billing process, these codes are used to determine medical necessity. Coders must make sure the procedure they are billing for makes sense with the diagnosis given. [11]

There are two types of ICD available ICD-9-CM & ICD-10-CM. The latest revisionICD-11 has more complicated coding structures and guidelines for medical coding.[2]

The **"CM"** at the end stands for **"clinical modification."** So the technical name for this code is the International Classification of Diseases, Ninth Revision, and Clinical Modification. The clinical modification is a set of revisions put in place by the National Center for Health

Statistics (NCHS), which is a division of the Center for Medicare and Medicaid Studies (CMS).

2. ICD-9 & ICD-9-CM

This code was initially intended for epidemiological purposes but has since become an integral part of the reimbursement cycle. While ICD codes are still used to track the incidence and spread of diseases and injury, their most important facet today is demonstrating medical necessity in claims.[12], [13]

ICD-9-CM codes are three-to-five digit numeric and, in certain cases, alpha-numeric codes. The first three digits in code are called the "category." The category describes the general illness, injury, or condition of the patient. In many cases, the category is not specific enough to describe the full extent of the patient's condition. Take dementia, for example. The basic ICD-9-CM code for dementia is 290.

In cases where more specificity is needed, a decimal point is added after the category and one or two more digits are added. The fourth digit of the ICD-9 code is called the "subcategory," and the fifth digit is called the "sub-classification".

The subcategory describes the etiology (cause), site, or manifestation of the disease or condition. The subdivision provides even more information about the site, cause, or manifestation of a disease, and is used only when the subcategory cannot provide sufficient information.

In medical coding, coders always want to code to the highest level of specificity. That means reading the report carefully, taking notes, abstracting the information, looking up the code, paying attention to all of the conventions for the code, and double-checking your work.[8]

Let's understand the ICD -9 codes with an example for better understanding**:**

2.1 CODE TREE EXAMPLE

123 – {Disease}

(*Category– First three digits)

123.0 – {Disease} in Chest

(*Subcategory –The digit after the decimal. It relates an important designation about the disease.)

123.00 -… uncomplicated

123.01 - … with complications in the cardiac system

123.02 - … with complications in the digestive system

(*Sub-classification–Last digit. This gives even more info about the designation outlined in the subcategory.)

123.0 – Disease in legs

And so on…

If we were to select 123.02 as our code, we'd read the full code as "Disease in the chest, with complications in the digestive system.[8]

"Most ICD-9-CM codes give the guidelines, or convention, which help to guide the coder to the accurate code for the diagnosis. These conventions may be punctuation or verbal instructions.

Table No. 1: CONVENTIONS IN ICD-9-CM [8]

CONVENTIONS	MEANING
Brackets [ ]	Enclose synonyms, alternative wordings, or explanatory phrases. Also used to identify manifestation codes
Parentheses ( )	Enclose supplementary words; whether absent or present, that nonetheless does not affect the code of the disease.
"Excludes"	Terms, conditions, or manifestations listed under an "Excludes" are coded elsewhere. In some cases, these "excluded" terms may not be coded with the code they are listed under
"Includes"	Comes immediately after the three-digit code. Further defines or gives examples of the term listed in the category.
"See"	This indicates that another term or code should be referenced instead of the listed code.

"See Also"	This indicates that another term or code may prove useful in the coding process. Unlike "See," "See Also" is not mandatory.
"Code First"	Indicates that the coder should list a particular code first. This typically happens with an underlying condition that has multiple manifestations, like diabetes. In situations like this, the underlying conditions are coded first, and then the manifestation is coded. "Code first" codes typically appear in the manifestation codes.
"Use Additional Code"	This phrase usually appears under the condition code (again, we'll use diabetes), and informs the coder that other codes for manifestations are available

Table No. 2: CONVENTIONS IN ICD-9-CM

CONVENTIONS	MEANING
"In Diseases Elsewhere Classified"	This note is attached exclusively to manifestation codes. It means that this manifestation is directly related to an underlying condition. A code with this note attached to it can never be used as the primary code (it could never have a "code first" note).
"Not Elsewhere Classified"	Abbreviated as "NEC," you may find this attached to a disease or condition that is not classified in the coding manual. Think of this as an unspecified code. An example of this might be category 995: "Certain adverse effects not elsewhere classified," which includes "anaphylactic reaction due to unspecified food" and other catch-all terms
"Not Otherwise Specified"	Abbreviated as "NOS," you'd turn to this in cases where the doctor or reporting physician has not provided

Table No. 3: CONVENTIONS IN ICD-9-CM

CHAPTER	CONCEPT	RANGE
1	Infectious and parasitic diseases	001–139
2	Neoplasm	140–239
3	Endocrine, nutritional and metabolic diseases, and immunity disorders	240–279
4	Diseases of the blood and blood-forming organs	280–289
5	Mental disorders	290–319
6	Diseases of the nervous system	320–359
7	Diseases of the sense organs	360–389
8	Diseases of the circulatory system	390–459
9	Diseases of the respiratory system	460–519
10	Diseases of the digestive system	520–579
11	Diseases of the genitourinary system	580–629
12	Complications of pregnancy, childbirth, and the puerperium	630–679
13	Diseases of the skin and subcutaneous tissue	680–709
14	Diseases of the musculoskeletal system and connective tissue	710–739
15	Congenital anomalies	740–759
16	Certain conditions originating in the prenatal period	760–779
17	Injury and poisoning	800-999
E-Codes	External causes of injury	e800–e999
V-Codes	Supplementary classification of factors influencing health status and contact with health services	V01–V91

2.2 Using ICD-9 CM:

The coding process begins with the medical report. The medical report provides the coder with an immense amount of information, including the patient's demographic info, their medical history, the patient's present symptoms, the doctor's diagnosis, the procedures performed by the doctor to ascertain or confirm the diagnosis, and the prescriptions or treatments, if any, recommended by the doctor. The medical report is the full documentation of the patient's visit.

After reading over the medical report, a coder will take notes and abstract the information in the report. Most physicians or providers will list the patient's symptoms and then give their diagnosis in a straightforward, direct manner.[14]

Let's assume,

A Patient who is a 28-year-old male. Had a height and weight 1.85m and 85 kg. Smoker. Had a History of asthma and breathing problems as a child, though none recently.

The patient presents with a cough, difficulty breathing, production of mucus, fever. The suspected diagnosis of acute bronchitis. A pulmonary function test was performed with a spirometer. Diagnosis confirmed.

Prescribed bed rest and a low dose of anti-inflammatory drugs (Prednisone) to patients.

There's a lot of information here! There's the patient's height, some of his medical history, his symptoms, the procedure the doctor performed, the prescription, and more. Since, however, we're looking at diagnosis codes, we have to winnow down what we're looking for.

In the case of a positive diagnosis, we don't code any symptoms. That means the only diagnostic code we're using is the one for acute bronchitis (466.0). That means you won't code for fever, hacking cough, or mucus production. You only code for symptoms when a healthcare provider is unable to make a clear, definitive diagnosis.

You might look at the report and also see "history of asthma and breathing problems," but since those conditions did not directly affect the patient's visit this time, we don't always have to code them.

In certain cases, you may find something missing from a medical report. This could be a procedure, or it could be an incomplete diagnosis. This can be especially difficult in medical reports on large, complicated procedures. There may also be multiple diagnoses listed in a medical report. A coder has to list every diagnosis (or set of symptoms) that's directly related to a procedure performed by the provider. Remember, ICD codes are used to demonstrate medical necessity in insurance claims. They justify the processes performed by the doctor. If you read a report and a certain procedure is not justified by a doctor's diagnosis, you must contact the doctor to get clarification.

The final step of the coding process is the submission of codes. In the past, this was done via paper forms, but today almost all medical codes are submitted via a software system.

Now you know what ICD-9-CM is, and how to use it, ICD-10-CM, which has replace ICD-9-CM in 2014. This code sets shares a lot of similarities but has a few critical differences.

2.3 Using ICD-10 CM:

ICD codes are updated every 10 to 15 years, and typically just include expansions to existing code sets. There is a bit of a challenge for incoming medical coders. The ICD-9-CM set is still in use today, and coders must be able to use it quickly and efficiently. But, the ICD-10-CM upgrade is looming, and coders must also be able to use that code set effectively.

As we mentioned, the ICD-10-CM code is similar to ICD-9-CM in what it does but distinct in its format and layout. Where ICD-9-CM is five digits and almost entirely numeric, ICD-10-CM is seven characters and entirely alphanumeric. As the field of medicine has grown by leaps and bounds in the years since ICD-9-CM was implemented, the code set has struggled to keep up. There are now too many new diseases, diagnoses, procedures for the code set to keep up.

We remember coding is all about the highest level of specificity," you can see that this is a real problem. As medical practice has developed and diversified, the old ICD-9-CM code set is increasingly incapable of providing the exact right codes.[14]

ICD-10-CM the code set is significantly larger than ICD-9-CM (there are 13,000 ICD-9 codes and 68,000 ICD-10 codes), and its increased number of subcategories and sub-classifications (the digits after the decimal point), allow for a far greater level of specificity in coding. ICD-10-CM is also more flexible and was designed in such a way that eases the entrance of codes for new, recently discovered, or expanded diagnoses.

ICD-10-CM is a seven-character, alphanumeric code. Each code begins with a letter. That letter is followed by two numbers. Like ICD-9-CM, the first three characters of ICD-10-CM are the "category." The category describes the general type of injury or disease. The category is followed by a decimal point and the subcategory. This is followed by up to two sub-classifications, which further explain the cause, manifestation, location, severity, and type of injury or disease. The last character is the extension.[11]

The extension describes the type of encounter this is. That is, if this is the first time a healthcare provider has seen the patient for this condition/injury/disease, it's listed as the "initial encounter." Every encounter after the first is listed as a "subsequent encounter." Patient visits related to the effects of a previous injury or disease are listed with the term "sequela".

The first digit of an ICD-10-CM code is always an alpha, the second digit is always numeric, and digits three through seven may be alpha or numeric. Here's a simplified look at.

2.4 ICD-10-CM's FORMAT

A01 – {Disease}

A01.0 {Disease] of the lungs

A01.01 … simple

A01.02 … complex

A01.020 … affecting the trachea

A01.021 … affecting the cardiopulmonary system

A01.021A … initial encounter

A01.021D … subsequent encounter

A01.021S … sequela

As above example, ICD-10-CM branches much farther out than ICD-9-CM. ICD-10-CM allows us to code the location and manifestation of a disease or injury far more accurately and the extensions reduce the administrative burden by documenting both the diagnosis and whether this injury or illness has been examined before.

Table No. 4: ICD-10-CM CODE MANUAL[8]

TOPIC	RANGE
Certain infections and parasitic diseases	A00-B99
Neoplasm	C00-D49
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D50-D89
Endocrine, nutritional and metabolic diseases	E00-E89
Mental, Behavioral and neurodevelopment disorders	F01-F99
Diseases of the nervous system	G00-G99
Diseases of the eye and adnexa	H00-H59
Diseases of the ear and mastoid process	H60-H95
Diseases of the circulatory system	I00-I99
Diseases of the respiratory system	J00-J99
Diseases of the digestive system	K00-K95
Diseases of the skin and subcutaneous tissue	L00-L99
Diseases of the musculoskeletal system and connective tissue	M00-M99
Diseases of the genitourinary system	N00-N99
Pregnancy, childbirth, and the puerperium	O00- O9A
Certain conditions originating in the prenatal period	P00-P96
Congenital malformations, deformations, and chromosomal abnormalities	Q00-Q99
Symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified	R00-R99
Injury, poisoning, and certain other consequences of external causes	S00-T88
External causes of morbidity	V00-Y99
Factors influencing health status and contact with health services	Z00-Z99

Table No. 5: ICD-9-CM VS. ICD-10-CM

ICD-9-CM	ICD-10-CM
Injury: Closed fracture of the the distal phalanx of a right index finger	Injury: Closed fracture of the the distal phalanx of a right index finger
Injury and Poisoning	S00-T88 - Injury, poisoning and certain other consequences of external causes
810-819 Fracture of Upper limb	S60-S69 – Injuries to the wrist, hand and fingers
816 Fracture of 1+ phalanxes of hand 816.0 Closed fracture of 1+ phalanges of the hand 816.00 ...unspecified 816.01 – closed fracture of middle or proximal 1+ phalanges of the hand 816.02 – closed fracture of distal 1+ phalanges of the hand	S62 Fracture at wrist and hand level S62.0 Fracture at the navicular [scaphoid] bone of the wrist S62.5 Fracture of thumb S62.6 Fracture of other and unspecified finger(s) S62.60 fracture of unspecified phalanx of finger S62.61 displaced fracture of proximal phalanx of finger S62.63 displaced fracture of distal phalanx of finger S62.630 displaced fracture of distal phalanx of an index finger S62.630A… initial encounter for closed fracture S62.630B … initial encounter for open fracture S62.630D … initial encounter for fracture with routine healing

ICD-10-CM and ICD-9-CM share more similarities as we have seen in the above example.[8]

ICD-10-CM allows coders to code to a much higher level of specificity. ICD-10-CM introduces laterality, to which side the injury or infection is on and substantially increases the amount of information about the diagnosis. Instead of leaving off at “closed fracture of distal

phalanx of hand," as we would in ICD-9-CM, we can go into fine detail about what type of fracture, on which finger, on which hand, and even we can visit this is for knowing the particular injury. [15]

EXTENSIONS & PLACEHOLDERS IN ICD -10CM

ICD-10-CM has another important convention that has to do with the code's extensions. Remember, extensions typically provide information that encounters this is for the healthcare provider with the patient. These are not always included, but in the case, if they are included, they cannot simply be appended to the end of whatever code is attached. Extensions are only found in the seventh character of an ICD-10-CM code. If a coder has to include an extension for an initial encounter on a code that does not have six characters, they must add placeholder characters. Coders use an 'X' for the placeholder digit.[11]

Example – A coder needs to code an instance of poisoning by unintentional under dosing of the antibiotic penicillin, the coder would use T36.0X1A. In this case, the fifth digit is empty, and so we'd use the placeholder character 'X.' remembers that placeholder characters are only used when an extension is necessary. Most ICD-10-CM codes do not include an extension for the encounter. [16]

3. Current Procedural Terminology (CPT)

Current Procedural Terminology (CPT). This expansive, important code set is published and maintained by the American Medical Association (AMA), and it is, with ICD, one of the most important code sets for medical coders to become familiar with.

CPT codes are used to describe tests, surgeries, evaluations, and any other medical procedure performed by a healthcare provider on a patient. This code system set is extremely large and includes the codes for thousands upon thousands of medical procedures.

Like ICD codes, CPT codes are also used to track important health data and measure performance and efficiency. Government agencies can use CPT codes to track the prevalence and value of certain procedures, and hospitals may use CPT codes to evaluate the efficiency and abilities of individuals or divisions within their facility.[11], [17]

Let's give a little closer, how these codes look like and are used. Each CPT code is five characters long and may be numeric or alphanumeric, depending on which category the CPT code is in. Now don't get confuse this with the 'category' in ICD. Remember that in ICD

codes the ‘category’ refers to the first three characters of the code, which describe the injury or disease documented by the healthcare provider.

With CPT, ‘Category’ refers to the division of the code set. CPT codes are divided into three categories. Category (I) is the most common and widely used set of codes within CPT. It describes most of the procedures performed by healthcare providers in inpatient and outpatient offices and hospitals. Category (II) codes are supplemental tracking codes used primarily for performance management. Category (III) codes are temporary codes that describe emerging and experimental technologies, services, and procedures.[8]

CPT codes have five digits, CPT is designed for flexibility and revision, and so there is often a lot of “space” between codes. Unlike ICD, each number in the CPT code does not correspond to a particular procedure or technology.[11]

Category (I) CPT codes are divided into six large sections based on which field of health care they directly pertain to.

The six sections of the CPT codebook are, in order:

- Evaluation and Management
- Anesthesiology
- Surgery
- Radiology
- Pathology and Laboratory
- Medicine

Each of these sections has its subdivisions, which correspond to what type of procedure, or what part of the body, that particular procedure relates to. Like ICD codes, many CPT codes are arranged by indentation. If a procedure is indented below another code, the indented procedure is an important or noteworthy variation on the above procedure and would replace the first code.

Certain codes have related procedures indented below them. These indented codes are important variations on the code above them and denote different methods, outcomes, or approaches to the same procedure. For example, the code for the elevation of a simple,

extradural depressed skull fracture is 62000. The code for the elevation of a compound or comminuted, extradural depressed skull fracture is 62005.

Recognize that the CPT code set has several instructions that inform the medical coder on how to best code the procedure performed. Remember that you always need to code to the highest level of specificity, and a miscoded procedure can be the difference between an accepted and rejected claim. The CPT code set also instructs coders on when to use multiple codes, when to use codes in tandem with one another (add-on codes), and which codes are "modifier exempt".

CPT Category (II) codes - These codes are five character-long, alphanumeric codes that provide additional information to the Category I codes. These codes are formatted to have four digits, followed by the character F. These codes are optional but can provide important information that can be used in performance management and future patient care. These codes never replace Category I or Category III codes, and instead, simply provide extra information. They are divided into numerical fields, each of which corresponds with a certain element of patient care.[7], [8]

3.1 CATEGORY II EXAMPLE

Composite Codes – These codes combine several procedures that typically occur in conjunction with one main procedure.

Example: 0001F: heart failure assessed (which includes all of them like)

- Blood pressure measured
- Level of activity assessed
- Clinical symptoms of volume overload assessed
- Weight recorded
- Clinical signs of volume overload assessed

1. Patient Management – Includes patient care provided for specific clinical purposes like pre-and postnatal care

Example: 0503F: Postpartum care visit

2. Patient History –Describes measures for select elements of patient history or symptom review

Example: 1030F: Pneumococcal immunization status assessed

3. Physical Examination

Example: 2014F: Mental status assessed

4. Diagnostic/Screening Processes or Results –Includes results of tests ordered, including clinical lab tests and radiological procedures

Example: 3006F: Chest X-ray documented and reviewed

5. Therapeutic, Preventive, or Other Interventions – Describes pharmacologic, procedural or behavioral therapies

Example: 4037F: influenza immunization ordered or administered

Category (III) of CPT codes - Is made up of temporary codes that represent emergent or experimental services, technology, and procedures. In certain cases, you may find that a newer procedure does not have a Category I code. There are codes in Category I for unlisted procedures, but if the procedure, technology, or service is listed in Category III, you are required to use the Category III code.

Category III codes allow for more specificity in coding, and they also help health facilities and government agencies track the efficacy of new, emergent medical techniques.

Like Category II, these codes are five characters long and are comprised of four digits and a terminal letter. In this case, the last letter of Category III codes is T. For example, the code for the fistulization of sclera for glaucoma, through the ciliary body is 0123T.

When a simple CPT code isn't enough, we turn to CPT modifiers. These important additions to CPT codes give extra information about how, where, and why a procedure was performed.

Since medical procedures and services are often complex, we sometimes need to supply additional information when we're coding. CPT Modifiers, like modifiers in the English language, provide additional information about the procedure. In English, a modifier may describe the, who, what, how, why, or where of a situation. Similarly, a CPT modifier may describe whether multiple procedures were performed, why that procedure was necessary,

where the procedure was performed on the body, how many surgeons worked on the patient, and lots of other information that may be critical to a claim's status with the insurance payer.

Certain modifiers may allow a healthcare provider to ask for more money from a payer. Modifiers -22 is one such modifier: If a surgeon performs a procedure that requires significantly more time to complete, due to a complication during the surgery, that procedure may be coded with a -22 at the end, for increased procedural services. Essentially, this modifier lets the payer know that the healthcare provider did more work than the basic CPT code would imply, and should be compensated for that work.[18]

CPT Modifiers are always two characters, and maybe numeric or alphanumeric. Most of the CPT modifiers you'll see are numeric, but there are a few alphanumeric. CPT modifiers are added to the end of a CPT code with a hyphen. In the case of more than one modifier, you code the "functional" modifier first, and the "informational" modifier second. The distinction between the two is simple: you always want to list the modifiers that most directly affect the reimbursement process first.[11], [18]

4. Healthcare Common Procedure Coding System (HCPCS)

HCPCS was developed by the Centers for Medicare and Medicaid (CMS) for the same reasons that the AMA developed CPT - for reporting medical procedures and services.

HCPCS codes to represent medical procedures to Medicare, Medicaid, and several other third-party payers. The code set is divided into three levels. Level one is identical to CPT, though technically those codes, when used to bill Medicare or Medicaid, are HCPCS codes. CMS looked at the established CPT codes and decided that they didn't need to improve upon or vary those codes, so instead, they folded all of CPT into HCPCS.[18]

Level II HCPCS codes are designed to represent non-physician services like ambulance rides, wheelchairs, walkers, other durable medical equipment, and other medical services that don't fit readily into Level I. Where CPT describes the procedure performed on the patient, it doesn't have many codes for the product used in the procedure. HCPCS Level II takes care of those products and pieces of medical equipment.

Level II codes are, like Level I, five characters long, but Level II codes are alphanumeric, with a letter occupying the first character of the code. These codes, like those in ICD and CPT, are grouped by the services they describe and are in numeric order.[11], [18]

4.1 HCPCS Level II CODE ALPHABETIC GROUPING:

Here's the full breakdown of HCPCS Level II codes by their alphabetic grouping:

A-codes: Transportation, Medical and Surgical Supplies, Miscellaneous and Experimental

B-codes: Enteral and Parenteral Therapy

C-codes: Temporary Hospital Outpatient Prospective Payment System

D-codes: Dental codes

E-codes: Durable Medical Equipment

G-codes: Temporary Procedures and Professional Services

H-codes: Rehabilitative Services

J-codes: Drugs administered other than oral method, chemotherapy drugs

K-codes: Temporary codes for durable medical equipment regional carriers

L-codes: Orthotic/prosthetic services

M-codes: Medical services

P-codes: Pathology and Laboratory

Q-codes: Temporary codes

R-codes: Diagnostic radiology services

S-codes: Private payer codes

T-codes: State Medicaid agency codes

V-codes: Vision/hearing service

HCPCS code manuals have an index and a large table of drugs. Whenever a coder is coding the delivery of a drug or medication, they should always use the drug table. Coding for medication is one of the most important parts of using HCPCS, and the drug table will provide much more accurate information on where to find the correct code.

Coders use HCPCS codes much like they would ICD or CPT codes. Upon receiving a medical report, you'd take notes on which procedure was performed, which products were prescribed, injected, or otherwise delivered to the patient, and then you'd use your HCPCS code set to find the appropriate code.[11], [18]

Be aware that when coding with HCPCS, you're going to have to strive for an even higher level of specificity than with CPT. Since this code set has codes for all different variations and amounts of equipment and medicine, you'll have to stay as close to the medical report as possible to make sure you're coding the correct procedure.

In the HCPCS manual, you'll recognize a lot of symbols from the CPT manual. Like CPT, HCPCS alerts you to which codes are new and which codes have been revised. New codes are listed with a circle, while revised codes have a triangle next to them. HCPCS is constantly being updated, and CMS, which maintains the code set, will often recycle codes. HCPCS features some strikethrough codes, and these let you know that a code that used to be listed there has been deleted and moved elsewhere.

Note that many codes in HCPCS Level II have specific guidelines for their use. Those guidelines are too various and fine grain to go over here, but you should know that with HCPCS, you always need to be paying attention. The diligent coder always takes note of the type of equipment used and the amount of medication delivered to the patient.

HCPCS modifiers allow for greater accuracy in coding and can be extremely important in the reimbursement process. HCPCS modifiers, like CPT modifiers, are always two characters and are added to the end of an HCPCS or CPT code with a hyphen. When differentiating between a CPT modifier and an HCPCS modifier, all there's one simple rule: if the modifier has a letter in it, it's an HCPCS modifier. If that modifier is entirely numeric, it's a CPT modifier.[8], [11], [18]

Table No. 6: HCPCS MODIFIERS IN CPT MANUAL

HCPCS MODIFIERS IN CPT MANUAL	
E1: upper left eyelid	LC: left circumflex coronary artery
E2: lower left eyelid	LD: left anterior descending coronary artery
E3: upper right eyelid	LT: left side (used to identify procedures performed on the left side of the body)
E4: lower right eyelid	
FA: left hand, thumb	
F1: left hand, second digit	QM: ambulance service provided under an arrangement by a provider of services.
F2: left hand, third digit	
F3: left hand, fourth digit	
F4: left hand, fifth digit	QN: ambulance service furnished directly by a provider of services
F5: right hand, thumb	
F6: right hand, second digit	
F7: right hand, third digit	RC: right coronary artery
F8: right hand, fourth digit	RT: right side (used to identify procedures performed on the right side of the body)
F9: right hand, fifth digit	
GG: performance and payment of a screening mammogram and diagnostic mammogram on the same patient, same day	TA: left foot, the great toe
	T1: left foot, second digit
	T2: left foot, third digit
	T3: left foot, fourth digit
GH: diagnostic mammogram converted from screening	T4: left foot, fifth digit
mammogram on same day	T5: right foot, the great toe
	T6: right foot, second digit
	T7: right foot, third digit
	T8: right foot, fourth digit
	T9: right foot, fifth digit

Certain HCPCS modifiers don't "agree" with certain CPT modifiers. The most obvious example of this would be CPT modifier -50 and the HCPCS modifiers –LT and –RT. These

modifiers are mutually exclusive: CPT modifier -50 describes a bilateral procedure, while HCPCS modifiers –LT and –RT describe which side of the body a procedure is performed on.

Working in medical coding sometimes requires finding equivalencies between different code sets. The code sets CPT, HCPCS, and ICD are updated annually, and medical coders need to know how to find and map codes that may have changed between updates.[11], [18]

5. CROSSWALKING OF CODING

We perform a task called cross-walking. The term 'cross-walking' actually comes from computer science. Put simply, cross-walking is the mapping of equivalent, identical, or similar information across two or more distinct data sets. Put another way, when you crosswalk codes, you perform a coding translation between two sets, not unlike how coders translate medical reports into codes in the first place.

We crosswalk between similar code sets or code sets that perform the same, or highly similar, functions. That is, the code sets we're translating between the need to both describe the same thing. Remember that cross-walking is not the process of finding the correct diagnosis code for a particular procedure.[11]

Most cross-walking is done between two versions of the same code set. That is a newer version and its older, now out-of-date version. For instance, the AMA updates the CPT code set every year, adding, changing the definitions or descriptions of, and deleting codes. In certain cases, you'd find older CPT codes. For the most part, CPT and HCPCS make this easy for you by listing deleted and updated codes in appendices in the back of each code manual.

The real cross-walking challenge for the medical coder is between ICD-9-CM and ICD-10-CM. ICD-10-CM; ICD-9-CM is out of date and no longer able to effectively represent new medical diagnoses. The latest revisionICD-11 has more complicated coding structures and guidelines for medical coding.[2]

ICD-9-CM has five characters and is primarily numeric, with a few alphanumeric codes used in certain situations. ICD-10-CM, on the other hand, is seven characters long and entirely alphanumeric. Where each ICD-9-CM code could have one subcategory and one sub-classification, and ICD-10-CM code can have one subcategory and two sub-classifications, in

addition to an alpha extension that provides information as to which visit, or encounter, this is with the patient's particular illness or injury.[18]

ICD-10-CM is a much more extensive, detail-oriented code set, but its new format and organization present coders with a challenge. ICD-10-CM format makes this cross-walking process difficult. Remember, as coders, we always to be as exact as possible.[16]

6. CODING SOFTWARE

In the past, coders entered their codes into paper forms, which they then passed on to the medical billing individual or organization. Today, to speed up the coding process and ensure more accuracy, the majority of the medical coding profession uses some type of coding software.

Software programs like Epic, Centricity, Advanced MD, Flash Code, Eclipse, and others have fields where coders can enter the correct procedure and diagnosis codes. These software programs may come with look-up tools that help coders find the correct code, but coders should always use their coding manuals to get the last word on which codes to use.

Epic is considered by many to be the gold standard and one of the more complicated programs. Coders who familiarize themselves with the Epic software program should be well suited to mastering other coding programs. These programs are often paired with medical billing programs.

The benefits of working with coding software are numerous. Coders (and billers) can track claims and easily call up old reports to check for efficiency and errors. Coding software is also excellent for tracking data over long periods and for performance management evaluations. Software like Epic and Eclipse can help coders keep track of their lag time for codes, allowing for better organization and a faster turnaround in the reimbursement cycle. Many coding programs also check, automatically, for compatible codes, though the responsible person is always on the coder to check and double-check using their coding manuals.[8]

Due to the complexity, a manual encoding task is time-consuming, error-prone, and expensive. Computer-aided coding systems are developed to recommend code candidates to human coders so that they can speed up manual encoding and avoid oversights for some rare diseases.

There are two main challenges for automatic medical code assignment. From the aspect of textual quality, medical records are not always structured in the same way. It is very difficult to extract important and relevant knowledge from various kinds of medical records effectively. Meanwhile, medical records are less formal than textual descriptions of ICD codes. Doctors usually write diagnosis in an informal and ungrammatical way, with telegraphic phrases, abbreviations, and typos.

The difficulties and the importance of ICD coding make the automatic system become an essential issue. Need to achieve a solution to provide medical coding assistance from un-structured textual data (medical reports, observations, clinical notes, etc.). The realization of an automated solution for medical coding support is of interest. This solution is based essentially on the combination of two approaches, namely - the NLP (Natural language processing) and the multi-class classification by machine learning. [10], [19]

Main Contributions are as follows:

- The realization of a workflow (pipeline) ensuring the preprocessing of the medical reports by exploiting the techniques of NLP.
- The use of machine learning techniques and algorithms to realize prediction models of the diagnoses and acts. This is the Naive Bayes and the Support Vector Machines (SVM) algorithms.

6.1 Medical coding support system

The automatic processing of medical information is based on the collection of clinical data relating to hospital stays of patients. As a result, considerable efforts have been made to construct the terminological resources used to express these data in coded form, similar to ICD-10. However, despite the increasing use of these terminologies, natural language remains the privileged vector of information and medical knowledge. From hospital reports to diagnostic and therapeutic protocols, the text is omnipresent. A medical coding support system is a system that analyzes medical documents and produces the appropriate codes according to the sentences and terms mentioned on these documents.

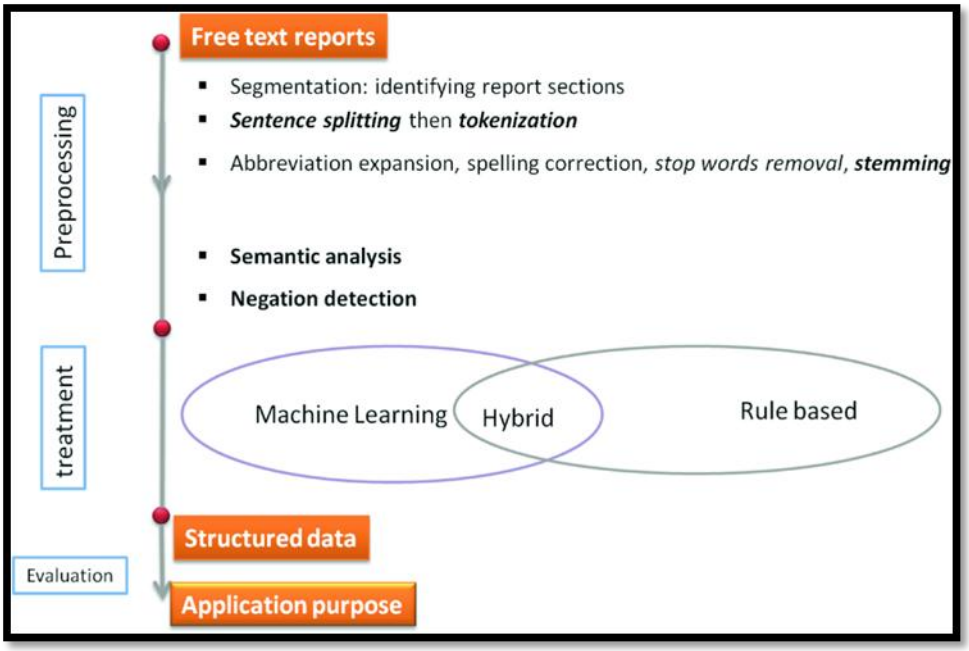

Figure No. 1: pipeline of a medical coding support system[19]

The pipeline of medical coding support system

Figure 1 [19] - illustrates the sequence of steps commonly used for the realization of a medical coding support system. In the first step, known as preprocessing, the contents of the reports are structured in sections. For each section, the text is subdivided into sentences (sentence splitting) and words (tokenization). At the word level, additional normalizations can be applied to obtain the lexical root of the term, referred to as stemming. This includes correcting misspellings and replacing abbreviations with their complete forms and eliminating stop words.

Secondly, the treatment step requires as input the result of the preprocessing step. It can be done according to three types of approaches, namely, a classification approach based on machine learning, a rule-based approach, or a hybrid approach combining the first two. An evaluation step of the coding support system succeeds immediately in the processing phase. Many studies have used NLP techniques associated with automatic classification approaches to extract coded clinical information from unstructured medical reports expressed in English.[20]

6.2 Automatic classification of textual medical reports

The classification consists of assigning instances of a given domain described by a set of discrete or continuous value attributes to a set of classes, which can be considered as values of a selected discrete target attribute. The process of classifying texts is based on a learning set D = [d_1, d_n] composed of documents.

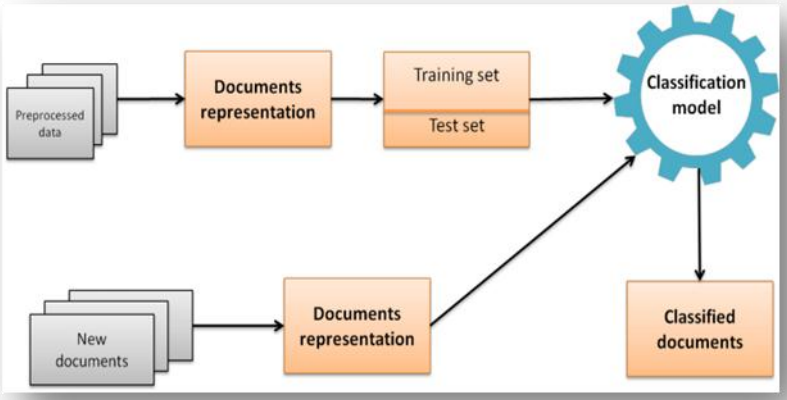

Figure No. 2: Process of automatic classification of documents[21]

Labeled with classes {C_1,C_k}.In the case of medical coding assistance, these classes may correspond to diagnosis codes expressed according to ICD-10. This enables to create a classification model based on a corpus of labeled document able to assign the correct classes to any new document d. This corpus is partitioned into two sets - one for learning and another for testing. As illustrated in Figure 2, the classification process consists first in training the model with the learning set. Once learned, its effectiveness will be tested with the test set.

It should be noted that before beginning the construction of the classification model, it is important to make the textual documents understandable by the learning algorithms. This occurs at the step designated by Document Representation illustrated in Figure 2.[21]

Representation of textual documents. The representation by the bag of words is often preferred. It consists to describe the content of a text utilizing descriptors (words or groups of words). The idea is to transform the different documents of a corpus into vectors where each element of a text vector represents textual units or simple words. The set of these words will be referred to as vocabulary. In the vector model, the components of a vector are determined according to the occurrences of words in the text.

Let C be a corpus of textual documents of size n and Di a document belonging to C. Let m be the number of terms and T = {$T_1 \ldots T_m$} be the set of these terms. In the vector representation, the document D_i is represented by a vector V_i. The collection of texts can be represented by a matrix whose columns represent the words and lines that represent the documents as illustrated in Figure 3. The Wij represent the weighting of a term T_j in the document D_i where ***0 < j < m, 0 < i < n.***

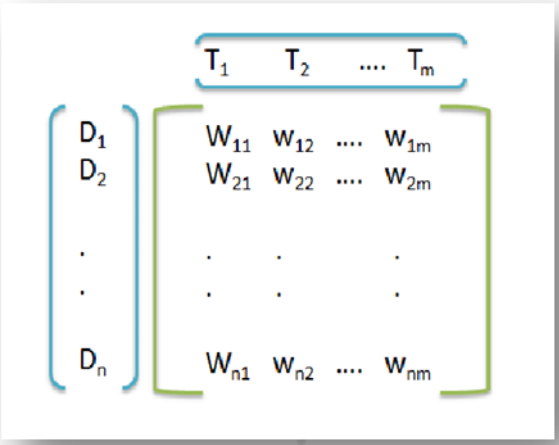

Figure No. 3: Documents terms matrix[22]

There are several possibilities for defining the weight of the words W_{ij}. Among others, Boolean weighting, weighting with word frequencies, TF IDF, etc.

Data description (documents-terms matrix) - The implementation of predictive techniques requires the mandatory passage through a phase of representation of the data. The nature of the predictive variable, which is in our case textual document, denotes the peculiarity of operation, which was achieved by using the bag of words representation to bring the description of the text corpus back to a table.

Individuals (documents) - Variables (terms): this is the documents-terms matrix described in Fig-3 [22]

The construction of the documents-terms matrix is done in two steps:

- First consists of a descriptor selection operation, designated by the generation of the vocabulary representing the corpus.
- A second step is designated by text formatting.

In this sense, we have considered words as descriptors and the Term Frequency (TF) × Inverse Document Frequency (IDF) as a type of text formatting, allowing gelatinizing the importance of a word in a document (TF) by its importance in the corpus (IDF). TF × IDF is defined as follows:

$$TF \times IDF\ (tk; Dj) = N\ (tk; DJ) * log\ \{|Tr|\ /\ T\ r(tk)\ \}$$

Where:

- N(tk; Dj): the number of occurrences of the descriptor tk in the document DJ ;
- |Tr|: the number of documents in the learning corpus;
- Tr (tk): the number of documents in the set in which the descriptor t_k appears once at least.

6.3 Machine Learning - Machine Learning is a computer science and statistics subdomain that aims to solve problems in different scientific fields. It is deeply related to artificial intelligence and optimization. It allows the creation of mathematical models from data. Many techniques and algorithms inherent in machine learning have been proposed in the literature. We present below two variants of these algorithms namely: Naive Bayes and SVM. The prediction models generated by these algorithms have proved their worth in the classification of documents.

- **Naive Bayes** -Naive Bayes is a probabilistic algorithm based on Bayes theorem. The naive aspect is due to the assumption that the variables are independent and fixed at the beginning. In the classification of texts, the descriptors are designated by variables.

The hypothesis of independence of the descriptors of the Naive Bayes model makes it simple and effective. Its training does not require many documents. This classification model has proven itself in the classification of short documents, including emails (Ham / Spam).

- **Support Vector Machines (SVM)** - is methods that come from an accurate and advanced mathematical analysis of the learning problem and based on binary separator in a vector space. The separators are hyperplane. To choose the best hyperplane, the idea of margin is solicited. The margin of a separating hyperplane is the smallest distance that separates it from the nearest.

Points. The SVM algorithm favors the hyperplane which ensures the largest possible margin. SVMs are powerful tools, which often obtain the best classification performance.[20]

Machine Learning. In the classification of texts by machine learning, it is important not to use the same data for modeling and evaluation of prediction models.

In this case, it is a question of dividing data into two parts: a first sample, called learning is used to elaborate the model; a second one, called test, is used to measure the performance. This task occurs before the construction of the matrix documents terms. Indeed, the texts of the test corpus must intervene neither in the constitution of vocabulary nor in the calculation of the weights of the matrix used for learning.

6.4 Experimental Approach

After preparing learning and test data, the construction of the document-term matrix was carried out according to the following two criteria:

- The abundance of descriptors in the corpus: it is the maximum number of documents where a descriptor appears. Indeed, the more a descriptor is common in the corpus, the less it will be used as discriminating between classes.
- The size of the vocabulary: it is the number of descriptors to be selected from the corpus.

The challenge is to find the combination of the two parameters above, which will lead to better system accuracy. For this purpose, an experimental study involving our data is required. It consists of varying the variable characterizing the abundance of descriptors in the corpus as well as the size of the vocabulary representing the latter.

For each combination of these variables, a prediction model is generated and is evaluated in terms of usual NLP based system performance indicators that are defined in Evaluation metrics (in section 6.5) (precision, F-measure, recall rate). Since our goal is to provide practitioners with coding support for medical information, evaluation metrics are calculated so that the correct code appears among the first k propositions of the system (in section 6.5).

6.5 Evaluation metrics

Evaluation is an important step in the same way as the other phases of the prediction model construction process. Several indicators reflect the success or the failure of a prediction

model. In our case, we define the evaluation commonly used in the literature for the evaluation of medical coding Support systems, in this case:

- Precision: The ratio of the number of documents correctly classified in a class to the number of documents to which this class is assigned.

- Recall: It is the ratio between the numbers of documents correctly classified in a class on the number of documents belonging to this class.

- F-Measure: It is an indicator that combines recall and precision. It is given by the following formula:

$$F - Measure = 2 * \frac{(Precision * Recall)}{(Precision + recall)}$$

7. CONCLUSION

The main purpose of the present paper is to propose a solution that reduces the burden of medical coding on practitioners. Precisely, we have developed a solution that helps hospital practitioners during medical coding. This solution affords a list of relevant diagnoses and acts codes that best match a given clinical situation. Doing so, we reduce the searching time spent by practitioners in the selection of required ICD-10 and GNPA codes. An empirical evaluation of the proposed solution with real clinical data provides preliminary evidence for the effectiveness of our proposal.[8]

Future work will focus on the following perspectives:

- Develop a solution to differentiate between different types of diagnosis (main, associated, etc.) based on multi-label classification. This evolution will best help the practitioner when choosing diagnoses codes.

- The performance improvement of the proposed solution, in terms of accuracy, is based on the use of structured data in electronic patient records (medications, laboratory results, etc).

REFERENCES:

[1] A. Tatham, Ed., 'The increasing importance of clinical coding', *Br. J. Hosp. Med.*, vol. 69, no. 7, pp. 372–373, Jul. 2008, DOI: 10.12968/hmed.2008.69.7.30409.

[2] F. Teng, Z. Ma, J. Chen, M. Xiao, and L. Huang, 'Automatic medical code assignment via a deep learning approach for intelligent healthcare', *IEEE J. Biomed. Health Inform.*, pp. 1–1, 2020, DOI: 10.1109/JBHI.2020.2996937.

[3] F. Ma *et al.*, 'Incorporating medical code descriptions for diagnosis prediction in healthcare', *BMC Med. Inform. Decis. Mak.*, vol. 19, no. S6, p. 267, Dec. 2019, DOI: 10.1186/s12911-019-0961-2.

[4] D. T. Heinze, 'LiveCode: A Deployed Application for Automated Medical Coding', p. 13.

[5] Y. Yan, G. Fung, J. G. Dy, and R. Rosales, 'Medical coding classification by leveraging inter-code relationships', in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, Washington, DC, USA, 2010, p. 193, DOI: 10.1145/1835804.1835831.

[6] I. A. Amarouche, D. Ahmed Zaid, and T. Kenaza, 'Implementation of a Medical Coding Support System by Combining Approaches: NLP and Machine Learning', in *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*, vol. 11470.

[7] 'mbacc_ebook_full_pages.pdf'.

[8] 'What is Medical Coding?' https://www.medicalbillingandcoding.org/learn-more-about-coding/ (accessed Aug. 22, 2020).

[9] Admin, 'Coder (Medical)', *Sisters' Action Network (SAN)*, Nov. 05, 2015. https://www.sanstudio.org/coder-medical/ (accessed Aug. 22, 2020).

[10] L. Cui, X. Xie, and Z. Shen, 'Prediction task guided representation learning of medical codes in EHR', *J. Biomed. Inform.*, vol. 84, pp. 1–10, Aug. 2018, DOI: 10.1016/j.jbi.2018.06.013.

[11] 'What is Medical Billing and Coding? - MBAC.org'. https://www.medicalbillingandcoding.org/what-is-mbac/ (accessed Aug. 22, 2020).

[12] M. Li *et al.*, 'Automated Icd-9 Coding Via a Deep Learning Approach', *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 16, no. 4, pp. 1193–1202, Jul. 2019, DOI: 10.1109/TCBB.2018.2817488.

[13] 'ICD Brief', *National DPP Coverage Toolkit*. https://coveragetoolkit.org/icd-brief/ (accessed Aug. 22, 2020).

[14] L. Zhang *et al.*, 'Automated grouping of medical codes via multiview banded spectral clustering', *J. Biomed. Inform.*, vol. 100, p. 103322, Dec. 2019, DOI: 10.1016/j.jbi.2019.103322.

[15] R. Tsopra *et al.*, 'The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care', *Int. J. Med. Inf.*, vol. 115, pp. 35–42, Jul. 2018, DOI: 10.1016/j.ijmedinf.2018.03.015.

[16] K. M. McGrew *et al.*, 'Validity of International Classification of Diseases codes in identifying illicit drug use target conditions using medical record data as a reference standard: A systematic review', *Drug Alcohol Depend.*, vol. 208, p. 107825, Mar. 2020, DOI: 10.1016/j.drugalcdep.2019.107825.

[17] K. A. Kelley, H. E. Hoops, L. Palmer, N. A. Cohen, and K. J. Brasel, 'Implementation of a medical coding curriculum for surgery residents', *Am. J. Surg.*, vol. 217, no. 5, pp. 834–838, May 2019, DOI: 10.1016/j.amjsurg.2019.02.027.

[18] S. Writers, 'Medical Billing Training - Expert Video Overview', *MedicalBillingandCoding.org*, Jul. 13, 2020. https://www.medicalbillingandcoding.org/billing-training/ (accessed Aug. 23, 2020).

[19] 'Fig. 1. The pipeline of a medical coding support system [4].', *ResearchGate*. https://www.researchgate.net/figure/Pipeline-of-medical-coding-support-system-4_fig1_326841378 (accessed Aug. 23, 2020).

[20] V. Gadepally, T. Mattson, M. Stonebraker, F. Wang, G. Luo, and G. Teodoro, Eds. Cham: Springer International Publishing, 2019, pp. 133–147.

[21] 'Fig. 2. Process of automatic classification of documents', *ResearchGate*. https://www.researchgate.net/figure/Process-of-automatic-classification-of-documents_fig2_326841378 (accessed Aug. 23, 2020).

[22] 'Fig. 3. Documentsterms matrix', *ResearchGate*. https://www.researchgate.net/figure/Documentsterms-matrix_fig3_326841378 (accessed Aug. 23, 2020).