CCS Coding of Discharge Diagnoses via Deep Neural Networks

Chadi Helwe
Computer Science
American University of Beirut
cth05@aub.edu.lb

Shady Elbassuoni
Computer Science
American University of Beirut
se58@aub.edu.lb

Mirabelle Geha
Emergency Medicine
American University of Beirut
mg67@aub.edu.lb

Eveline Hitti
Emergency Medicine
American University of Beirut
eh16@aub.edu.lb

Carla Makhlouf Obermeyer
Epidemiology & Population Health
American University of Beirut
cm39@aub.edu.lb

ABSTRACT

A standard procedure in the medical domain is to code discharge diagnoses into a set of manageable categories known as the CCS codes. This is typically done by first manually coding the discharge diagnoses into the standard ICD codes and then using a one-to-one mapping between ICD and CCS codes. In this paper, we study the applicability of deep learning to perform automatic coding of discharge diagnoses into CCS codes. In particular, we build an LSTM network combined with a dense neural network that uses medically-trained word embeddings to code discharge diagnoses into single-level CCS codes. We also investigate the advantage of mapping discharge diagnoses into UMLS concepts before coding is carried out. Experimental results based on a large dataset of manually coded discharge diagnoses show that our deep-learning model outperforms the state-of-the-art automatic coding approaches and that the mapping to UMLS concepts consistently results in significant improvement in the coding accuracy.

CCS CONCEPTS

•Computing methodologies →Supervised learning by classification; Classification and regression trees; Neural networks;
•Applied computing →Health informatics;

KEYWORDS

deep learning, CCS coding, discharge diagnosis, ICD coding

1 INTRODUCTION

Discharge diagnoses are short pieces of free-text that describe the final diagnoses provided to patients by health-care professionals upon their release from hospitals. Discharge diagnoses of patients are typically coded into a set of standard codes, known as the International Classification of Diseases (ICD) codes [1] for billing purposes and for statistical analysis and reporting. However, the process of coding discharge diagnoses is carried out manually by expert medical coders, which makes the coding process very costly and also subject to human errors.

Moreover, discharge diagnoses are usually further coded into a limited number of broad diagnosis categories known as the Clinical Classification Software (CCS) codes [2]. CCS collapses the over 14,000 ICD codes into 285 mutually exclusive categories, known as the single-level CCS codes. To be able to code a discharge diagnosis into a single-level CCS code, the diagnosis must be first *manually* coded into ICD and then the CCS tool can be used to obtain the required CCS codes. Note that without coding discharge diagnoses into CCS codes, the large number of ICD codes makes statistical analysis and reporting difficult and time consuming [6]. Table 1 shows some example discharge diagnoses and their corresponding ICD and single-level CCS code descriptions.

In this paper, we investigate the applicability of deep-learning to automate this coding process. Particularly, we aim to bypass the stage of ICD coding and *directly* map discharge diagnoses into single-level CCS codes. As mentioned earlier, the process of coding diagnoses into ICD codes is a tedious process that requires a lot of manual effort. Some attempts have been carried out to automate the process of ICD coding, however this problem has been shown to be very challenging for various reasons. First, the large number of possible ICD codes makes it very difficult to build robust models that can accurately predict these codes. In fact, most previous work either focused only on certain types of medical records such as radiology reports [7, 8, 11], or restricted the prediction to a limited number of ICD codes only [14, 24]. Second, ICD coding involves looking at detailed medical reports and in some cases laboratory test results, etc. This makes predicting ICD codes based only on discharge diagnoses practically impossible in most cases, as shown by the inconsistent and mostly poor results obtained by most previous work. On the other hand, the relatively small number of single-level CCS codes provides promise that it might be sufficient to deduce these codes based only on discharge diagnoses.

In addition to the difficulty of automatically predicting ICD codes, it is also a very sensitive process. ICD coding is a key source of data for billing purposes and is used to monitor population health and inform policy decisions. Thus, any errors that might occur due to automating this process, which is inevitable, could have drastic consequences, for instance healthcare fraud [17]. This makes the

[1] http://www.who.int/classifications/icd/en/

[2] https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

DH '17, July 2–5, 2017, London, United Kingdom
 ISBN 978-1-4503-5249-9/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3079452.3079498

Diagnosis	ICD Code Description	CCS Code Description
aortic valve disease / coronary artery bypass graft with avr /sda	mitral valve insufficiency and aortic valve stenosis	heart valve disorders
benzodiazepine overdose	poisoning by other sedatives and hypnotics	poisoning by other medications and drugs
left wrist laceration	open wound of wrist, with tendon involvement	open wounds
acute mental status changes/respiratory distress	delirium due to conditions classified elsewhere	delirium dementia and amnestic and other cognitive disorders
complete heart block	atrioventricular block, complete	conduction disorders

Table 1: Example diagnoses and their corresponding ICD and CCS code descriptions

medical community reluctant to use automated tools to perform ICD coding.

Another reason for directly obtaining CCS codes from discharge diagnoses is that in many scenarios, we are truly just interested in the CCS codes and have no use at all for ICD codes. For instance, in many developing countries, billing is not coupled or based on ICD codes. However, medical experts are constantly interested in performing data analytics on discharge data. As mentioned earlier, the large number of possible ICD codes makes statistical analysis and reporting rather meaningless. Thus, most medical experts rely on the CCS codes instead for data analytics. On the other hand, obtaining ICD codes manually might not be an option due to the lack of resources or expertise to perform the coding and automating this is not feasible either as explained earlier.

Given all of the above, we propose to directly map discharge diagnoses to CCS codes. To this end, we build a long short-term memory (LSTM) network equipped with a dense neural network (DNN) that classifies discharge diagnoses into single-level CCS codes. As mentioned earlier, these discharge diagnoses are typically very short pieces of free-text (i.e., few words), and thus relying on classical bag-of-words classification approaches will not be adequate. On the other hand, our deep-learning model utilizes word representations (i.e., embeddings) learnt from a large corpus of medical documents to represent diagnoses. By making use of word embeddings, we alleviate the issue of the short length of the diagnoses and take into consideration syntactic and semantic similarities between words when performing the coding.

Finally, we investigate the use of MetaMap [2], a tool that automatically maps biomedical texts into the Unified Medical Language System (UMLS) concepts. By doing this, we manage to unfold medical abbreviations and correct spelling mistakes in the diagnoses. Once this mapping is carried out, our classification task can be seamlessly carried out on the mapped data rather than the raw data. In other words, this mapping process can be seen as a preprocessing step.

2 RELATED WORK

Automating the coding process of medical records has been a constant goal in the domain of health informatics. All previous approaches focused on automating the ICD coding of medical records. For instance, Lita et al. [15] investigated machine learning models to classify patient medical records into ICD-9 codes. Goldstein and Arzumtsyan [11] proposed three approaches for automatic coding of radiology reports to ICD-9 codes. Simialry, Farkas and Szarvas [8] studied the problem of ICD-9 coding, limited to the case of radiology reports. Perotte et al. [19] experimented with two ICD-9 coding approaches. Perotte with another set of authors [20] introduced hierarchically supervised latent Dirichlet allocation and used it to assign ICD-9 codes to clinical records. Yan et al. [24] introduced a multilabel large-margin classifier that automatically learns the underlying inter-code structure and allows the controlled incorporation of prior knowledge about medical code relationships. Rios and Kavuluru [21] studied the role of feature selection, training data selection, and probabilistic threshold optimization in improving different multilabel classification approaches to extract ICD codes from EMRs. Finally, Subotin and Davis [23] studied a related problem, which is the prediction of ICD-10-PCS codes (procedural coding). To the best of our knowledge, no previous work has attempted to tackle the problem of *directly* coding medical records into CCS codes. Moreover, we could not find any studies reporting on the accuracy of obtaining CCS codes from automatically ICD-coded medical records either.

MetaMap and UMLS thesaurus have been used in different research projects in the biomedical domain. For example, Sanchez-Cisneros et al. [22] proposed a system for drug named entity recognition using dictionary-based and ontology-based methods. Lana-Serrano et al. [13] developed a rule-based approach using semantic information to recognize chemical compounds and drugs names. Aronson and Rindflesch [3] proposed the use of the MetaMap tool to associate a given medical query with UMLS concepts. The closest to our work is the work by Goldstein and Arzumtsyan [11] where the authors used MetaMap to expand the medical texts and used the expanded text to predict ICD-9 codes. To the best of our knowledge, no previous work had attempted to use UMLS resources such as the MetaMap tool to tackle the problem of CCS coding of medical records directly from discharge diagnoses.

Deep learning is starting to gain popularity in the biomedical domain, for instance for medical imaging, drug discovery, or for anomaly detection. For example, Cernazanu-Glavan and Holban [4] proposed a segmentation method of bone structure in x-ray images based on a convolutional neural network. Al Rahhal et al. [1] proposed a deep learning method for active classification of electrocardiogram signals. Lipton et al. [14] presented a study to empirically evaluate the ability of long short-term memory networks (LSTMs) to recognize patterns in multivariate time series of

	diagnosis	ICD Code	CCS Code
Number	14,720	2,789	241
Average	4.003	1.84	1.44
St. Div.	70.75	6.55	2.76

Table 2: Dataset Overview

clinical measurements. Choi et al [5] showed how to learn word embeddings of a wide range of concepts in medicine, including ICD-9 codes, medications, procedures, and laboratory tests. Our work is the first of its kind to utilize deep learning and word representations for the problem of automatic coding of discharge diagnoses.

3 METHODOLOGY

3.1 Dataset

Our dataset is based on the MIMIC-III dataset [12]. It consists of discharge diagnoses for over 58,000 deidentified patients and their manually assigned ICD-9-CM codes (9th revision of ICD - Clinical Modification). Each discharge diagnosis is mapped to one or more ICD codes. In addition, the same discharge diagnosis might be assigned to different ICD codes for different patients. In our dataset, the average length of a discharge diagnosis is 4.255 words and the standard deviation is 2.57.

To obtain the single-level CCS codes for each discharge diagnosis in our dataset, we relied on the CCS tool. The tool provides a file that contains a direct one-to-one mapping between each ICD code and the single-level CCS code. Thus, for a given discharge diagnosis, we use all its ICD codes and for each ICD code that is assigned to it, we retrieve the corresponding single-level CCS code. This means that we might end up with more than one possible single-level CCS code for each diagnosis in our dataset.

Table 2 summarizes our dataset. The first row displays the number of unique diagnoses, unique ICD codes, and unique single-level CCS codes in the whole dataset. The second row provides the average frequency of a diagnosis and the average number of ICD and CCS codes per diagnosis and the third row displays the standard deviations of these averages. As can be seen from the table, each diagnosis is repeated on average around 4 times and is associated with less than 2 ICD and CCS codes on average. Moreover, our dataset is somehow imbalanced with some ICD codes (and consequently CCS codes) dominating the dataset.

The dataset summarized in Table 2 is the ground-truth we will build our classifier on. Since each diagnosis might appear more than once in the dataset and might be associated with one or more CCS codes, we use a *majority vote* to obtain a single CCS code per diagnosis. That is, for each diagnosis in our dataset, we compute the frequency of each CCS code associated with it, and assign the most frequent code to the diagnosis. Our goal will then be to predict the most probable code for a given diagnosis. Alternatively, we could keep multiple codes for each diagnosis and develop a multilabel classifier to predict more than one code per diagnosis. However, it seemed more reasonable to us to evaluate the validity of our hypotheses on a simpler instance of the problem first. Recall that our three hypotheses were 1) whether predicting CCS codes directly from discharge diagnoses is better than predicting ICD codes first then using the CCS tool to obtain the CCS codes, 2) whether using deep-learning is better than using bag-of-words classical approaches, and 3) whether preprocessing using MetaMap is beneficial in the coding process. In future work, we will extend our model to handle the case of multilabel coding.

Diagnosis	MetaMap Concepts
interior myocardial infarction	cardiac infarction
necrotizing fascititis	error
hypertension	hypertensive disease
chiari malformation sda	error
uti	urinary tract infections

Table 3: Example Diagnoses and their corresponding MetaMap concepts

	diagnosis	ICD Code	CCS Code
Number	4,742	2,789	241
Average	12.42	3.82	2.37
St. Div.	137.33	15.50	5.99

Table 4: MetaMap Dataset

3.2 MetaMap Preprocessing

As mentioned earlier, discharge diagnoses are typically very short pieces of free-text or notes written by medical professionals. As a result, many of these diagnoses tend to have medical abbreviations and spelling mistakes. To be able to perform robust coding, we propose to use MetaMap [2], a tool that automatically maps biomedical texts into the Unified Medical Language System (UMLS) concepts, to preprocess the data. MetaMap uses a knowledge-intensive approach based on symbolic, natural-language processing and computational-linguistic techniques.

We preprocess our dataset using MetaMap as follows. We pass each diagnosis to MetaMap as input and retrieve a set of Concept Unique Identifiers (CUIs), where each CUI represents a UMLS concept. We then use the *preferred* names of the concepts identified and these represent the new instances which we base classification on. Table 3 shows some example diagnoses and the mapped-to concepts after preprocessing using MetaMap. After preprocessing, the average length of a discharge diagnosis is 3.31 words and the standard deviation is 1.77.

After mapping every diagnosis in our dataset, we end up with a new cleaned version of the dataset which is summarized in Table 4. Note that the unique number of diagnoses after preprocessing with MetaMap drops drastically since many different diagnoses would end up being mapped to the same string after preprocessing with MetaMap. Also note that for some diagnoses, MetaMap could not identify any concepts. For those diagnoses that were not mapped to any MetaMap concept, we just use the original diagnosis as is. Overall, out of the 14,720 unique diagnoses we had, 11,000 were mapped to 1,022 unique concepts and 3,720 did not produce any matches when they were preprocessed using MetaMap.

3.3 CCS Coding

We first describe how we obtained the word representations or embeddings that are used as input to our deep neural-network classification model. We then describe the general neural-network architecture we used for building our classifier.

To learn the word embeddings, we trained the Continuous Bag of Words (CBOW) model originally proposed by Mikolov et al. [16] over a medical corpus. Our medical corpus was constructed as follows. First, we retrieved the articles of all the diseases described in Wikipedia. Second, we wrote a web-scraping script to retrieve additional web pages describing diseases from medical websites like MedlinePlus, MayoClinic, John Hopkins Medicine, and National Health Service.

This whole process described above provided us with a medical corpus consisting of 10,403 documents. We then used the CBOW model on this corpus to generate the word embeddings. More specifically, each word was represented as a 300 dimensional, continuous and real-valued vector also known as word embedding.

The model that we used for our classification task, namely single-level CCS coding, is composed of a long short-term memory (LSTM) network followed by a dense neural network (DNN) with three hidden layers. First, we used our word embeddings to retrieve the embedding vector of every word. We then used an LSTM network to compute a vector for each diagnosis from the vectors of words it contains. An LSTM network is a kind of a recurrent neural network (RNN) developed by Hochreiter and Schmidhuber [10], and is capable of mapping vectors of words with variable length to a fixed-length vector by recursively transforming the current word vector with the output vector of the previous step.

We extended our LSTM network to create a deeper architecture by adding a dense neural network of three hidden layers and a $softmax$ activation function to predict the final labels of diagnoses. Adding a fully connected neural network on top of an LSTM will disentangle the factors of variations in the hidden state, making it easier to predict the output. It allows the hidden state of the model to be more compact and results in the model being able to summarize the history of previous inputs more efficiently [18]. Note that we also retrain our embedding matrix as part of the model parameters to be learnt.

4 EXPERIMENTS

We make three different hypotheses in this paper. The first hypothesis is that predicting the CCS codes directly from discharge diagnoses is more effective than predicting ICD codes first then using the CCS tool. To validate this, we compare our approach to a baseline approach that first predicts ICD codes and then uses the CCS tool to obtain the CCS codes from the predicted ICD codes. This baseline approach was based on the same deep neural network architecture used by our approach, with the only difference being that the goal is to predict ICD codes rather than CCS codes. That is, to be able to predict the ICD code from a given discharge diagnosis, we trained an LSTM network equipped with a dense neural network as explained in Section 3.3 and used the obtained classifier. To generate the training data, we again used a majority vote over the different ICD codes associated with a diagnosis to obtain a single ICD code per diagnosis.

Model	Raw Data	MetaMap Data
DNN Direct	0.863	0.959
DNN ICD + CCS Tool	0.788	0.845
SVM Direct	0.838	0.954
SVM ICD + CCS Tool	0.778	0.843

Table 5: F-measure of CCS coding for the different approaches

Our second hypothesis is that using deep learning based on word embeddings is more effective than using a classical coding approach that relies on bag-of-words. To validate this hypothesis, we compare our deep learning approach to a baseline Support Vector Machines (SVM) approach [9]. The SVM approach uses unigrams (i.e., words in the diagnosis) as features. We also build another version of our first baseline approach that predicts ICD codes using an SVM classifier and then use the CCS tool to obtain the single-level CCS codes from the predicted ICD codes.

Our third and final hypothesis was that preprocessing discharge diagnoses using the MetaMap tool would improve the coding process. To test this, we build four different versions of the above approaches based on the preprocessed data and compare them to the approaches using the raw data.

All our experiments were conducted on the MIMIC-III dataset [12], described in Section 3.1. The dataset was split into three folds as follows: 80% training, 10% validation, and 10% testing. All experiments were run on a Windows 10 machine with a 16 GB RAM, a CPU Intel Core I7 and a GPU NVIDIA GeForce GTX 1060 6GB.

Our model and the first baseline approach were trained on the training data using 100 epochs of stochastic gradient descent with batch size 5 and the training took 4.5 hours on average. The validation set was used to tune the parameters of the models such as the number of epochs, the initial pretrained embeddings (the medically trained ones or Glove) and the number of hidden layers of the dense neural network. The test set was used to evaluate the performance of the chosen model after the parameters were tuned on the validation set.

The SVM approach was trained in a similar fashion to the above two approaches. That is, we used the same 80% of the data to train the model using different hyperparameters such as different kernels, the value of the regularization parameter and so on. The validation set was then used to tune the hyperparameters in order to select the best model, which happened to be a linear soft-margin SVM. Finally, the chosen model was used to predict the CCS codes for the test data.

Table 5 displays the F-measure of our approach versus the baseline approaches on the test data. As can be seen, our approach that directly predicts the CCS codes (DNN Direct) outperforms the baseline approach that first predicts the ICD codes and then uses the CCS tool (DNN ICD + CCS Tool) by over 9.5% in F-measure.

Moreover, when we compare our approach to the baseline when both are trained on the preprocessed data (i.e., MetaMap data in Table 5), we see significant improvements in the accuracy of the prediction for both approaches, with over 11% improvement in F-measure in the case of our approach and over 7% in the case of the baseline (DNN ICD + CCS tool). This clearly highlights the benefit of using MetaMap to preprocess the discharge diagnoses before coding is carried out, regardless of whether we predict ICD codes first or directly predict the CCS codes. Moreover, our approach that directly predicts CCS codes still outperforms the ICD coding approach even on the preprocessed data with over 13% improvement in F-measure.

We also observe from Table 5 that our approach outperforms the second baseline approach, the SVM approach that directly predicts CCS codes using bag-of-words (i.e., unigrams). This clearly highlights the effectiveness of our deep learning model in predicting CCS codes. Moreover, the SVM approach that directly predicts the CCS codes from discharge diagnoses also outperforms the SVM approach that first predicts the ICD codes then uses the CCS tool to obtain the CCS codes (i.e., SVM ICD + CCS Tool). This confirms that our hypothesis that predicting CCS codes directly is better than predicting ICD codes first and then using the CCS tool. Finally, it is also noticeable that the SVM approaches, whether direct or used to predict ICD codes first, witness consistent improvements when using the MetaMap tool to first preprocess the data.

5 CONCLUSION

We proposed an approach to automatically code medical records such as discharge diagnoses. Our approach utilizes an LSTM network followed by a dense neural network to directly predict single-level CCS codes of discharge diagnoses. Our approach makes use of MetaMap to unfold abbreviations and correct spelling mistakes by mapping diagnoses to UMLS concepts. It also makes use of medically-trained word embeddings to represent diagnoses in a low-dimensional space. We compared our approach to three baselines. The first used our deep neural network architecture but predicted ICD codes first then used the CCS tool to obtain the CCS codes. The second was an SVM approach that uses unigrams to directly predict CCS codes. The final baseline was also an SVM approach that first predicts ICD codes then uses the CCS tool to obtain the CCS codes. Our experiments showed that our approach outperforms all baselines in terms of F-measure. Moreover, preprocessing with MetaMap was shown to be highly effective in improving the accuracy of all approaches.

6 COMPETING INTERESTS

The authors have declared that no competing interests exist.

7 HUMAN AND ANIMAL RIGHTS

The project has been approved by the American University of Beirut Institutional Review Board (IRB).

ACKNOWLEDGMENTS

The authors would like to thank the American University of Beirut Research Board (URB) for funding this project.

REFERENCES

[1] MM Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and RR Yager. 2016. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences* 345 (2016), 340–354.
[2] Alan R Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program.. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 17.
[3] Alan R Aronson and Thomas C Rindflesch. 1997. Query expansion using the UMLS Metathesaurus.. In *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, 485.
[4] Cosmin Cernazanu-Glavan and Stefan Holban. 2013. Segmentation of bone structure in X-ray images using convolutional neural network. *Adv. Electr. Comput. Eng* 13, 1 (2013), 87–94.
[5] Youngduck Choi, Chill Yi-I Chiu, and David Sontag. 2016. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings* 2016 (2016), 41.
[6] Healthcare Cost, Utilization Project (HCUP), et al. 2008. Clinical Classifications Software (CCS) for ICD-9-CM. 2008. (2008).
[7] Koby Crammer, Mark Dredze, Kuzman Ganchev, Partha Pratim Talukdar, and Steven Carroll. 2007. Automatic code assignment to medical text. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. Association for Computational Linguistics, 129–136.
[8] Richárd Farkas and György Szarvas. 2008. Automatic construction of rule-based ICD-9-CM coding systems. *BMC bioinformatics* 9, 3 (2008), S10.
[9] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. 1998. Support vector machines. *IEEE Intelligent Systems and their Applications* 13, 4 (1998), 18–28.
[10] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[11] MBA Ira Goldstein and MLS Anna Arzumtsyan. 2007. Three approaches to automatic assignment of ICD-9-CM codes to radiology reports. (2007).
[12] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3 (2016).
[13] Sara Lana-Serrano, Daniel Sanchez-Cisneros, Leonardo Campillos, and Isabel Segura-Bedmar. 2013. Recognizing chemical compounds and drugs: a rule-based approach using semantic information. In *BioCreative Challenge Evaluation Workshop*, Vol. 2. Citeseer, 121.
[14] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzell. 2015. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677* (2015).
[15] Lucian Vlad Lita, Shipeng Yu, Radu Stefan Niculescu, and Jinbo Bi. 2008. Large scale diagnostic code classification for medical patient records.. In *IJCNLP*. Citeseer, 877–882.
[16] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[17] William C Morris, Daniel T Heinze, Homer R Warner Jr, Aron Primack, AE Morsch, Ronald E Sheffer, Mark A Jennings, Mark L Morsch, and Michelle A Jimmink. 2000. Assessing the accuracy of an automated coding system in emergency medicine.. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 595.
[18] Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2013. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026* (2013).
[19] Adler Perotte, Rimma Pivovarov, Karthik Natarajan, Nicole Weiskopf, Frank Wood, and Noémie Elhadad. 2014. Diagnosis code assignment: models and evaluation metrics. *Journal of the American Medical Informatics Association* 21, 2 (2014), 231–237.
[20] Adler J Perotte, Frank Wood, Noemie Elhadad, and Nicholas Bartlett. 2011. Hierarchically supervised latent Dirichlet allocation. In *Advances in Neural Information Processing Systems*. 2609–2617.
[21] Anthony Rios and Ramakanth Kavuluru. 2013. Supervised extraction of diagnosis codes from EMRs: role of feature selection, data selection, and probabilistic thresholding. In *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 66–73.
[22] Daniel Sanchez-Cisneros, Paloma Martínez, and Isabel Segura-Bedmar. 2013. Combining dictionaries and ontologies for drug name recognition in biomedical texts. In *Proceedings of the 7th international workshop on Data and text mining in biomedical informatics*. ACM, 27–30.
[23] Michael Subotin and Anthony R Davis. 2014. A system for predicting ICD-10-PCS codes from electronic health records. In *Proc BioNLP*. Citeseer, 59–67.
[24] Yan Yan, Glenn Fung, Jennifer G Dy, and Romer Rosales. 2010. Medical coding classification by leveraging inter-code relationships. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 193–202.