

## Article

# Applying Deep Learning Model to Predict Diagnosis Code of Medical Records

Jakir Hossain Bhuiyan Masud <sup>1</sup>, Chen-Cheng Kuo <sup>1</sup>, Chih-Yang Yeh <sup>1</sup>, Hsuan-Chia Yang <sup>1,2,3,\*,†</sup>   
and Ming-Chin Lin <sup>1,4,5,\*,†</sup> 

- <sup>1</sup> Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan; d610105005@tmu.edu.tw (J.H.B.M.); m610105009@tmu.edu.tw (C.-C.K.); d610104001@tmu.edu.tw (C.-Y.Y.)
- <sup>2</sup> International Center for Health Information Technology (ICHIT), College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan
- <sup>3</sup> Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei 11031, Taiwan
- <sup>4</sup> Department of Neurosurgery, Shuang Ho Hospital, Taipei Medical University, New Taipei City 23561, Taiwan
- <sup>5</sup> Taipei Neuroscience Institute, Taipei Medical University, Taipei 11031, Taiwan
- \* Correspondence: lovejog@tmu.edu.tw (H.-C.Y.); arbiter@tmu.edu.tw (M.-C.L.)
- † These authors contributed equally to this work.

**Abstract:** The International Classification of Diseases (ICD) code is a diagnostic classification standard that is frequently used as a referencing system in healthcare and insurance. However, it takes time and effort to find and use the right diagnosis code based on a patient's medical records. In response, deep learning (DL) methods have been developed to assist physicians in the ICD coding process. Our findings propose a deep learning model that utilized clinical notes from medical records to predict ICD-10 codes. Our research used text-based medical data from the outpatient department (OPD) of a university hospital from January to December 2016. The dataset used clinical notes from five departments, and a total of 21,953 medical records were collected. Clinical notes consisted of a subjective component, objective component, assessment, plan (SOAP) notes, diagnosis code, and a drug list. The dataset was divided into two groups: 90% for training and 10% for test cases. We applied natural language processing (NLP) technique (word embedding, Word2Vector) to process the data. A deep learning-based convolutional neural network (CNN) model was created based on the information presented above. Three metrics (precision, recall, and F-score) were used to calculate the achievement of the deep learning CNN model. Clinically acceptable results were achieved through the deep learning model for five departments (precision: 0.53–0.96; recall: 0.85–0.99; and F-score: 0.65–0.98). With a precision of 0.95, a recall of 0.99, and an F-score of 0.98, the deep learning model performed the best in the department of cardiology. Our proposed CNN model significantly improved the prediction performance for an automated ICD-10 code prediction system based on prior clinical information. This CNN model could reduce the laborious task of manual coding and could assist physicians in making a better diagnosis.

**Keywords:** clinical note; natural language processing; deep learning; convolutional neural network; diagnosis code



**Citation:** Masud, J.H.B.; Kuo, C.-C.; Yeh, C.-Y.; Yang, H.-C.; Lin, M.-C. Applying Deep Learning Model to Predict Diagnosis Code of Medical Records. *Diagnostics* **2023**, *13*, 2297. <https://doi.org/10.3390/diagnostics13132297>

Academic Editor: Sameer Antani

Received: 8 May 2023

Revised: 23 June 2023

Accepted: 4 July 2023

Published: 6 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The World Health Organization (WHO) has published and maintains a list of diagnostic classifications called the International Classification of Diseases (ICD), which gives each illness a unique code [1,2]. ICD is a system of categorization and coding for disease. The ICD-10 version consists of more than 70,000 codes [3]. ICD codes have been broadly adopted by healthcare providers for healthcare reimbursement and retrieving diagnostic information [4,5]. Medical coders need to extract key information and assign correct codes based on categories within an electronic medical record (EMR) [6]. The complex structure

and amount of information in EMRs significantly increase the difficulty of manual coding. In the outpatient departments of hospitals, physicians manually assign ICD codes based on the information given in clinical notes; however, this is challenging, time-consuming, and prone to error [7]. Billing issues and underpayments can result from coding errors [8].

Automatic coding systems have become more popular as a result of recent developments in the fields of deep learning (DL) and natural language processing (NLP) systems. DL has shown promising outcomes in text classification [9–11]. EMR is a great source of text data as it includes clinical notes and discharge notes. Clinical notes, which include SOAP (subjective component, objective component, assessment, and plan) notes, drug lists, and ICD-10 codes, are a crucial resource for determining the nature of health issues. In recent years, automatic ICD coding has become a significant area of clinical medicine study. As a result, there is a huge need for an automatic ICD coding system.

Recent research has focused on methods linked to deep learning. The deep learning model can produce astounding outcomes in computer vision [12]. Zhang et al. [13] revealed a gated recurrent unit (GRU) network to predict medication on the basis of the disease codes. Wang et al. [14] used natural language processing to classify ICD-10 cm codes from hospital text data and achieved an F1-score of 0.62. Chen et al. [15] proposed a deep neural network (DNN) model to predict the ICD-10 clinical modification code and achieved an F1-score of 0.715. Wang et al. [16] revealed word2vec to predict ICD-10 cm code and achieved an F-score of 0.67.

The most popular deep learning architecture for explaining issues with natural language processing is the convolutional neural network (CNN) [17,18]. In order to categorize sentences using word2vec, Kim [19] presented a convolutional neural network model. A deep learning strategy was suggested by Gangavarapu et al. [20] to categorize the ICD-9 code in the nursing notes. A CNN model based on deep learning was utilized by Chen et al. [21] to predict the incidence of cerebral infarction from hospital data. The deep learning-based CNN techniques of Moons et al. [22] promoted the use of the discharge report for ICD-9 code classification.

Nursing notes are gathered using a multi-label classification system to anticipate ICD-9 [20]. Nursing notes are used in numerous studies to predict ICD-9 codes [23]. An app was created by Mauch et al. [24] to predict incisional hernias. Discharge summaries were used by Kavuluru et al. [25] to predict the EMRs' ICD codes. A deep learning algorithm was utilized by Li et al. [26] to forecast bone disease using EHR. Based on EHR data, Jin et al. [27] created a deep learning model to predict cardiac disease.

A research paper deals with the NLP methods to predict medical specialties from the unstructured text notes of a university hospital [28]. Vinod, P., et al. demonstrated a deep learning model from clinical text data [29]. Teng, F., et al. developed a deep learning model to predict ICD codes from free text data [30].

In the traditional machine learning approach for text classification, various processes are employed for feature extraction, such as TF-IDF, one-hot encoding, counts, and so forth. Deep learning, on the other hand, serves as a universal tool for optimizing the selection of these features. In this study, we demonstrate that the application of deep learning can yield significant improvements in text classification. There are many synonyms, acronyms, and typos in EMRs. Based on our experience, the vocabulary used in each department or by each doctor is different. If using rule-based methods, it will need to create a set of “rules” for each doctor.

Automatic ICD-10 coding has been the subject of numerous investigations. The model achievement of these approaches to automatically predicting ICD-10 codes has limitations, though, and it is not clinically sensible. By developing disease prediction models, healthcare professionals can gain insights to develop treatment plans and enhance the overall quality of patient care. The lack of clinically satisfactory studies motivates us to conduct this research. This prediction can identify the appropriate diagnosis codes with a probability score from text data. Then, the physician can identify proper diagnosis codes efficiently. This model can help in the completeness (predict missing diagnosis from EHR) of diagnosis.

We focused on a multi-label text categorization system for medical records in this work. The best ICD-10 codes were determined using a “top 10” ranking approach with a likelihood score for each prediction. Additionally, for each prediction, we manually checked the missing ICD-10 codes.

The use of a convolutional neural network (CNN) based multi-label text classification, with data taken from SOAP notes and drug lists, to predict ICD-10 codes is a new approach in this field. The objective of this work was to build a deep learning model that can assist doctors in choosing the most pertinent ICD-10 codes.

## 2. Methods

**Dataset:** This is a retrospective cohort study of patients. We collected clinical notes from the EMRs of a university hospital in Taipei, Taiwan. The dataset consists of clinical notes from the outpatient department (OPD) in the year 2016 (from January to December). The clinical notes consisted of SOAP (subjective component, objective component, assessment, and plan) notes, drug lists, and diagnosis lists (ICD-10 code). In the study, we focused on the three variables of SOAP notes, drug lists, and ICD-10 codes. The total number of clinical records was 21,953, which spanned five departments (Neurology, Psychiatry, Nephrology, Cardiology, and Metabolism). A total of 20,173 (90%) records were used for dataset training, while 1780 (10%) were used as a test set. In this study design (Figure 1), we selected 21,953 clinical notes from five departments to develop a deep learning-based CNN model. The Taipei Medical University Joint Institutional Review Board (TMU-JIRB) approved this study.

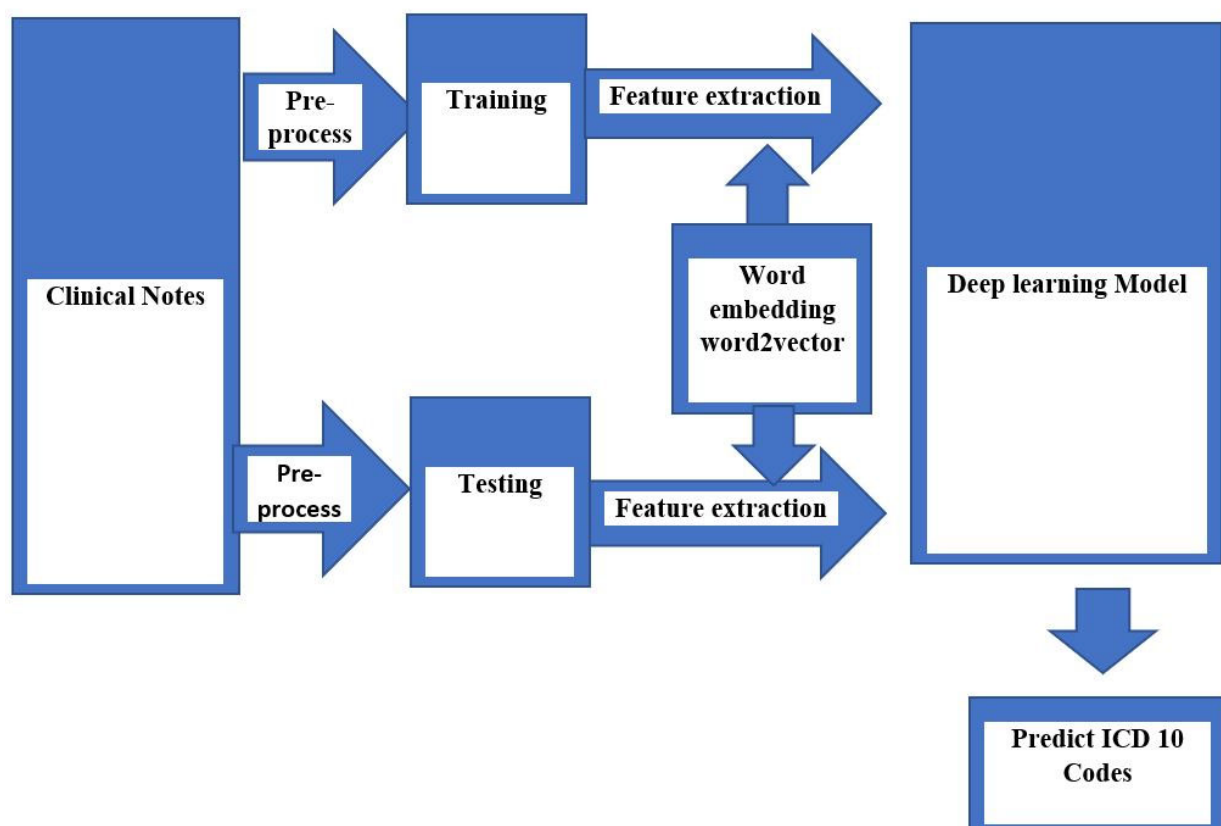


Figure 1. Study Design.

**Data Pre-processing:** Clean and relevant information was desired for developing a deep learning model. Punctuation, supplemental spaces, infrequent words, stop words, and redundant components were all eliminated. The Natural Language Toolkit package (version 3.8) and the Python package (version 3.8) were used to perform pre-processing before tokenizing the text [31,32].

**Extraction of feature:** In this study, we extracted features from the raw data using the word2vec method [33], and then we trained deep learning models using those features. Word embedding is a form of word representation that maps words onto real-number vectors by representing words in a vector space with many dimensions.

Models that create word embeddings are part of word2vec. Both the Continuous Bag of Words (CBOW) and the skip-gram designs are used by word2vec. While skip-gram attempts to predict several context words from a single input word, CBOW aims to predict a single word from a defined window size of context words.

The skip-gram model transforms a one-hot vector for each word based on a corpus of text data. The one-hot vector is used to convert a word into a vector made up entirely of 0; one coordinate, which represents the string, is equal to 1. A neural network with a single hidden layer receives the one-hot vector input. To change a distributed representation of words and employ a vector with a various number of dimensions, a sequence of text is used. Then, each word is carried out at random with different weight distributions among the components. The vector size is equal to the number of distinct words in a text (Figure 2).

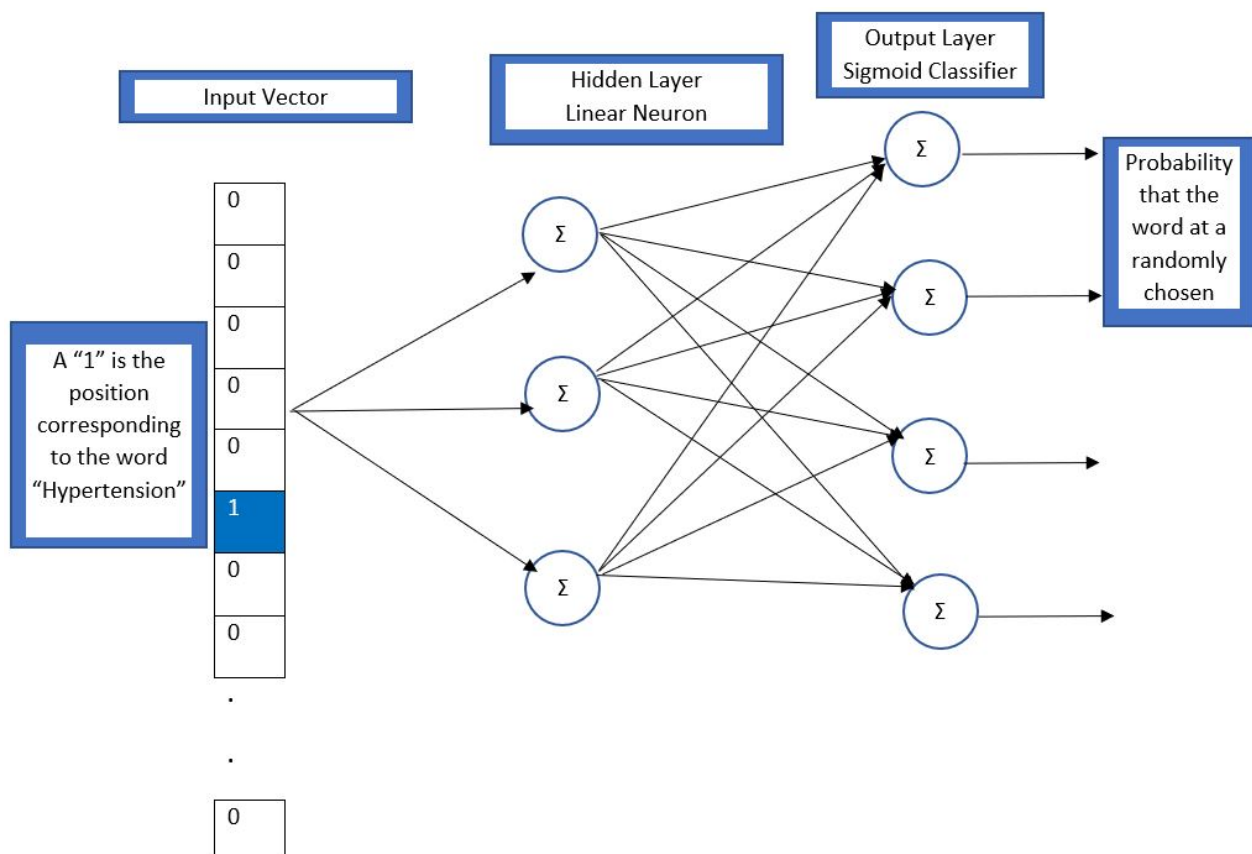
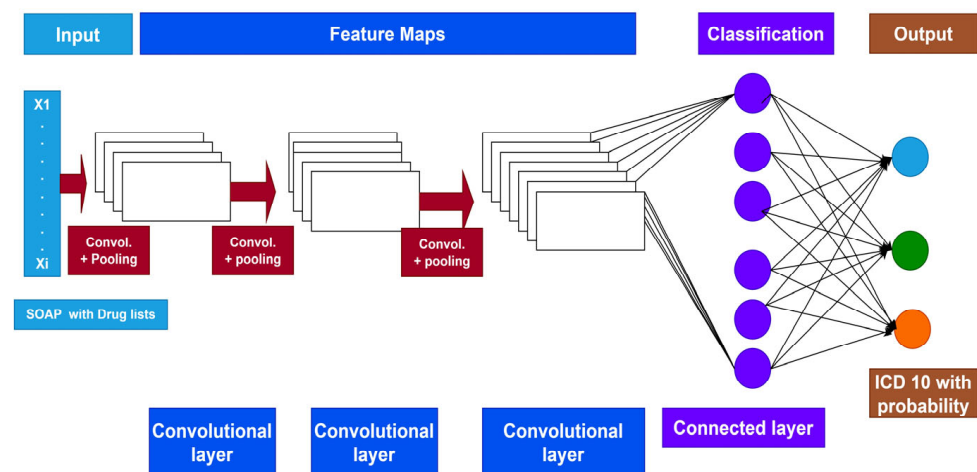


Figure 2. Architecture of neural network.

The input layer has a dimension of  $1 \times V$ , where  $V$  is the number of words in the corpus vocabulary (i.e., one-hot depiction of the word). The input layer is transformed into the hidden layer using the weight matrix. This hidden layer has a dimension of  $1 \times E$ , where  $E$  is the selected size of the word embedding. Finally, the weight matrix transforms the hidden layer into the output layer. In the hidden layer, we employed the sigmoid function. The last layer has a size of  $1 \times V$ , and each value in the vector represents the likelihood score of the target word in that position. We now have a weight matrix  $W$  of dimension  $V \times E$ , following the training of the entire vocabulary. This matrix links the input layer to the hidden layer. The weight matrices are then adjusted as the model gains experience in predicting the main word.

**Development of Deep Learning Model:** In the current study, we predicted ICD-10 codes using a CNN classification model based on deep learning. First, word2vec was used to create dense, low-dimensional feature vectors from the input words. We set the parameter fixed by sample length = 200 Matrix length and the vector dimension created by word2vec to 128 in the training. To capture various n-gram characteristics, the filter window sizes in the convolutional layer were 1, 2, 3, 4 and 5 with a fixed filter window width of 128. The most important features were then retrieved from the feature map for classification in the maximum pooling layer (Figure 3).



**Figure 3.** Deep learning CNN architecture.

The SOAP notes and drug lists used as inputs and the ICD-10 codes as output were used in this study to test the assignment as a multi-label text classification issue. Python and Keras [34] were utilized to carry out the word embedding word2vec CNN work for the multi-label text classification challenge. The final layer of CNN that we used was the sigmoid activation function. ICD-10 codes were predicted using a ranking (top 10) approach. This ranking system is able to categorize the best diagnoses in chronological order. For the training set, 90% of the data were used, and for the test set, 10% was utilized. Figure 4 depicts the study's overall design.

**Performance measure:** The deep learning-based CNN model performance was measured by three metrics (precision, recall, and an F-score).

**Precision:** This denotes the proportion of actual positive results to all positive results. Equation (1):

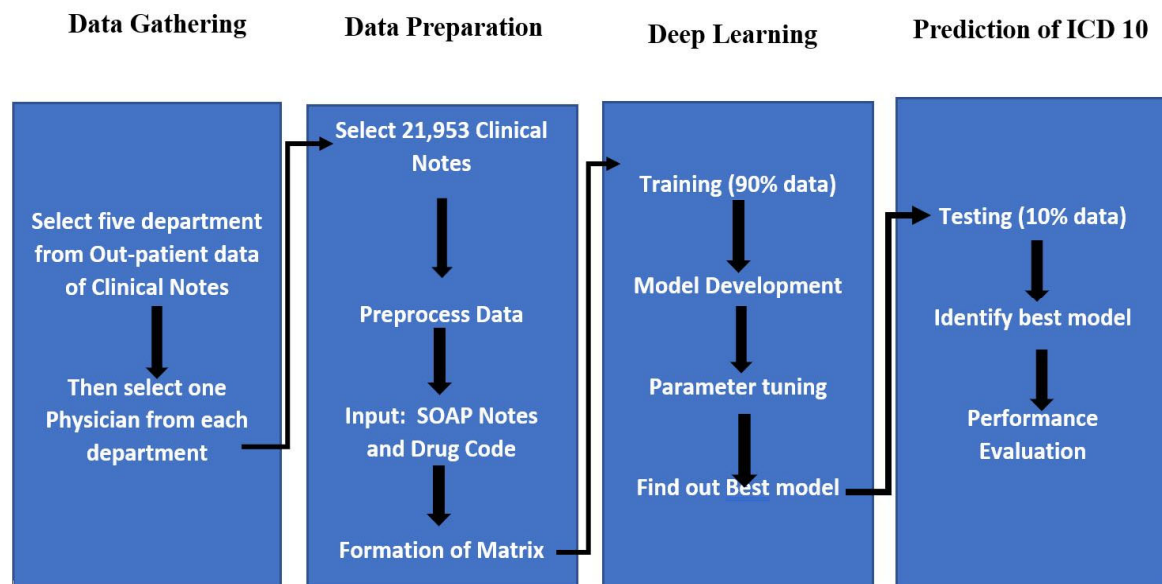
$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (1)$$

**Recall:** The recall denotes the number of positive results made out of all positive results in the dataset. Equation (2):

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (2)$$

**F-score:** The harmonic mean of recall and precision is an F-score. Equation (3):

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$



**Figure 4.** The overall layout of the study.

### 3. Results

**Characteristics of Data:** We retrospectively collected the data of all patients who visited a university hospital between January and December 2016. A total of 21,953 clinical notes were included in this study (Table 1). In our dataset, the majority of records (6027) were from the neurology department. Then, the psychiatry department provided 5789 records; the nephrology department provided 3707 records; the cardiology department provided 3668 records, and the metabolism department provided 2762 records in the total dataset. The age range of patients was between 10–101 years.

**Table 1.** Characteristics of data.

Characteristics	Frequency (n)
Total number of clinical notes from Doctors of five Departments	21,953
Doctors' notes from Neurology	6027
Doctors' notes from Psychiatry	5789
Doctors' notes from Nephrology	3707
Doctors' notes from Cardiology	3668
Doctors' notes from Metabolism	2762
Number of ICD 10 codes	1131
Number of Drugs	807

**Performance of Model:** Precision, recall, and an F-score were used to gauge how well the deep learning-based CNN model performed. With a precision of 0.96, a recall of 0.99, and an F-score of 0.98, the Department of Cardiology performed the best, followed by the Departments of Metabolism (F-score of 0.86), Psychiatry (F-score of 0.75), and Neurology (F-score of 0.71) (Table 2). Batch 64 produced a good model performance with a precision



of 0.66, a recall of 0.94, an F-score of 0.77, and an accuracy of 0.94 (Table 3). Accuracy in training and testing is 94% using 70/30 data (Table 4).

**Table 2.** CNN model performance for the doctor from five departments.

Name of Department	Test Cases	Number of ICD-10 Codes	Number of Drugs	Precision	Recall	F-Score	Accuracy
Doctors' notes from Cardiology	284	148	145	0.96	0.99	0.98	0.99
Doctors' notes from Metabolism	307	155	136	0.78	0.97	0.86	0.97
Doctors' notes from Psychiatry	475	193	128	0.64	0.91	0.75	0.91
Doctors' notes from Neurology	282	358	177	0.60	0.85	0.71	0.85
Doctors' notes from Nephrology	432	277	221	0.52	0.88	0.65	0.88

**Table 3.** Model performance using different batch sizes for one doctor.

Batch	Accuracy	Precision	Recall	F-Score	Loss	Error Rate	Computational Time (Minutes)
64	0.94	0.66	0.94	0.77	0.03	0.06	14
128	0.89	0.45	0.89	0.60	0.05	0.11	14
512	0.88	0.44	0.88	0.59	0.06	0.12	14
1000	0.84	0.40	0.84	0.54	0.07	0.16	14
1024	0.84	0.39	0.84	0.53	0.07	0.16	14
2000	0.81	0.38	0.81	0.52	0.08	0.19	14
2048	0.81	0.36	0.81	0.50	0.08	0.19	14

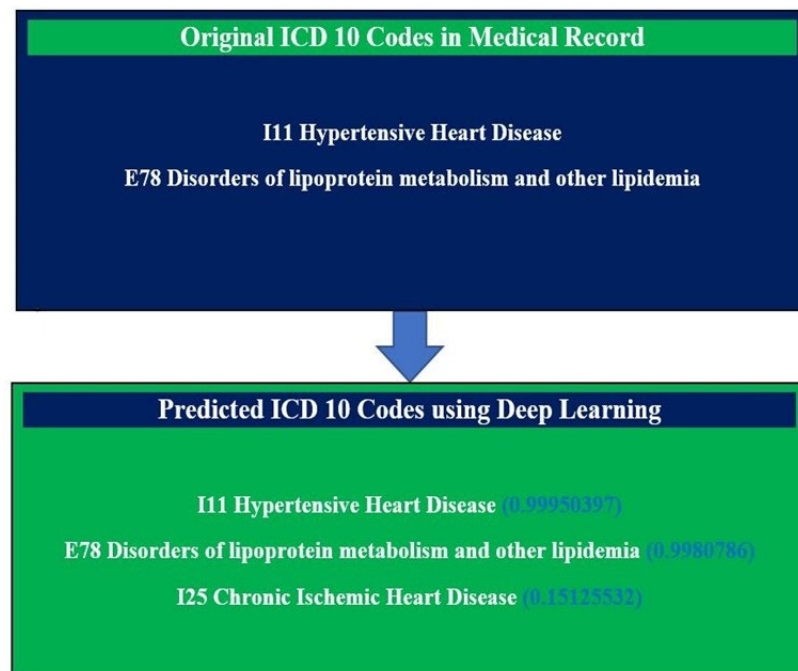
**Table 4.** Training and testing accuracy of the one doctor model.

Training Accuracy	Testing Accuracy
0.94	0.94

#### Evaluation:

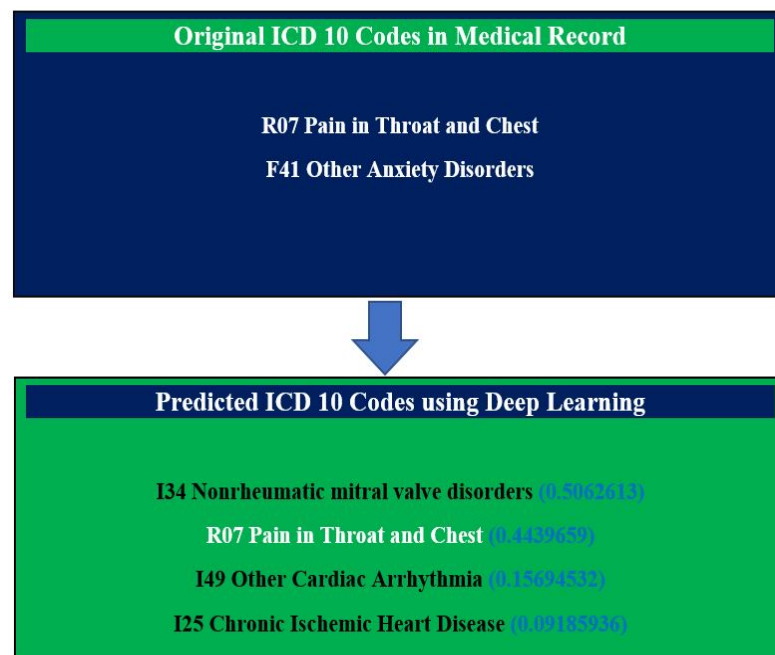
We evaluated our deep learning-based CNN model to identify its overall accuracy and appropriateness; a manual review was also conducted. Our CNN model appropriately predicts the necessary ICD-10 codes.

We further investigated the predicted ICD-10 codes and found that the CNN model has a high likelihood of accurate ICD-10 code prediction based on the words. For example, in the clinical notes for Figure 5, there were certain words used for chronic ischemic disease; our model was able to identify these words and correctly report a missing ICD-10 code. In the original medical records, the physician entered two diagnosis codes (ICD-10 codes), four drug codes, and SOAP notes for a patient. However, the CNN model predicted the appropriate disease codes (with one extra ICD-10 code) with a probability score based on clinical notes (Figure 5).



**Figure 5.** Evaluation of the performance of deep learning to predict ICD 10 codes with a probability score.

For the data in Figure 6, certain words were found in the clinical notes for cardiac arrhythmia, nonrheumatic mitral valve disorder, and chronic ischemic heart disease; the CNN model has a high chance of accurate ICD-10 code prediction based on the words and drug history. In the original clinical notes, the physician input two ICD-10 codes, five drug codes, and a SOAP note for a patient. However, the CNN model predicted the appropriate disease codes (three extra ICD-10 codes) with probability scores based on clinical notes (Figure 6). Thus, our model predicted the missing ICD-10 codes that could help the physician in better decision making.



**Figure 6.** Evaluation of the performance of deep learning to predict ICD 10 codes with probability scores for missing diagnoses.



#### 4. Discussion

**Main Findings:** Our study was designed to develop a deep learning-based CNN model to identify diagnosis codes automatically from clinical notes of medical records. This model achieved satisfactory performance in predicting ICD-10 codes using SOAP notes and drug lists. The performance of the deep learning model was the highest in the Department of Cardiology (with a precision of 0.95, a recall of 0.99, and an F-score of 0.98), followed by the Department of Metabolism (with a precision of 0.78, a recall of 0.97, and an F-score of 0.86), the Department of Psychiatry (with a precision of 0.64, a recall of 0.91, and an F-score of 0.75), and the Department of Neurology (with a precision of 0.60, a recall of 0.85, and an F-score of 0.71).

##### Compared to Previous Research:

A deep learning model was used in our study to forecast ICD-10 codes from clinical notes. By identifying relevant ICD-10 codes based on clinical notes, our CNN model has the enormous potential to improve the accuracy of ICD-10 coding and decrease the amount of human coding. The model performance in this study can correctly predict missing ICD-10 codes and is clinically satisfactory. Prior research has been conducted to evaluate the potential of deep learning-based models for ICD-10 code prediction. These deep learning models did not, however, perform in a clinically desirable manner.

A Micro F1 score of 63.42 was obtained for 50 distinct ICD-9 block codes in a study by Moons et al. [22] using a deep learning-based CNN model to predict ICD-9. The ICD-9 codes were categorized in their study using discharge summaries. A hierarchical deep learning model was created by Shi et al. [35] to predict ICD codes from discharge notes, and this model received an F-score of 0.53 for 50 ICD codes.

Suo et al. [36] employed convolutional neural networks and their model had an accuracy of up to 0.74 in predicting diabetes mellitus, obesity, and chronic obstructive pulmonary disease. A convolutional neural network model was utilized by Cheng et al. [37] to forecast the future recurrence of chronic heart failure and chronic obstructive pulmonary disease.

A light gradient-boosting machine (LightGBM) was employed by Diao et al. [38] to automate ICD-10 categorization from discharge summaries, and their best model produced a macro-averaged F1 (Macro-F1) score of 88.3%. Wang et al. [16] proposed a deep learning model to predict ICD-10 clinical modification codes from EHR data and their model achieved an F-score of 0.67.

A deep learning model was put forth by Rashidian et al. [39] to predict ICD-9 codes from EHR data. In this study, they predicted diabetes, acute renal failure, and chronic kidney disease using data from demographics, lab findings, and prescription information. Their deep learning model received an F1-score of 80.04 for diabetes, an F1-score of 66.86 for acute renal failure, and an F1-score of 75.77 for chronic kidney disease.

A deep learning method to predict ICD-19 codes from discharge summaries was put forward by Li et al. [40]. On the MIMIC-II dataset, their model obtained a micro F-measure of 0.335, while on the MIMIC-III dataset, it obtained a micro F-measure of 0.408. A deep learning model was also put forward by Choi et al. [41] to predict diagnosis codes using EHR data. A recurrent neural network model was applied in their study, which had a 79% recall rate.

In their study of Word2Vec convolutional neural networks for ICD-9 coding prediction, Hsu et al. [42] reported a micro F1 score of 0.76 for 19 different ICD-9 chapter codes from the discharge summary. The accuracy for 19 different ICD-9 chapter codes was 0.833, according to Gangavarapu et al. [43], who utilized a deep learning model trained on nursing notes. In order to resolve the multi-class labeling and multi-label classification technique, Samonte et al. [44] used an enhanced hierarchical attention network (EnHAN) and the word embedding method. Their model had an accuracy of 0.841.

A strategy to predict ICD-9 codes using clinical notes was put forward by Obeid et al. [45], and their model achieved an F-score of 0.769. From a subjective aspect of clinical notes, Hsu et al. [46] suggested a deep learning model and obtained an accuracy of 0.409 for 2017 different ICD-9 codes. For 2833 ICD-9 codes, Xie et al. [47] employed a deep

learning model trained on the diagnosis description and attained a sensitivity of 0.29. From a subjective feature of clinical notes, Singaravelan et al. [48] developed a deep learning model and attained a recall of 0.57 for 1871 ICD-9 codes.

Zeng et al. [49] developed a deep learning model to predict ICD-9 codes from discharge summaries and obtained an F1 score of 0.42 for 6984 ICD-9 codes. The application of deep learning to predict ICD-9 codes by Huang et al. [50] resulted in an F1 score of 0.6957 for 10 ICD-9 codes. In their investigation, ICD-9 codes were predicted using clinical notes.

Our study uses 1131 ICD-10 codes in the process of ICD-10 prediction. In this study, we predicted ICD-10 codes using SOAP notes and drug lists. This comparison demonstrates the originality of our research. We reviewed (Table 5) the deep learning model performance of work on ICD code prediction. It shows that our work is superior to previous investigations. The table compares the results of our investigation to those of earlier studies.

**Table 5.** Comparative evaluation of different studies.

Work	Data	Method	Target Variable	Performance Measure
Hsu et al. [42]	Discharge summary	Deep learning	(i) 19 distinct ICD-9 chapter codes, (ii) top 50 ICD-9 codes, (iii) top 100 ICD-9 codes	(i) Micro F1 score of 0.76, (ii) Micro F1 score of 0.57, (iii) Micro F1 score of 0.51
Gangavarapu et al. [43]	Nursing notes	Deep learning	19 distinct ICD-9 chapter codes	Accuracy of 0.833
Samonte et al. [44]	Discharge summary	Deep learning	10 distinct ICD-9 codes	Precision of 0.780, Recall of 0.620, F1 score of 0.678
Obeid et al. [45]	Clinical notes	Deep learning	ICD-9 code from E950-E959	Area under the ROC curve score of 0.882, F-score of 0.769
Hsu et al. [46]	Subjective component	Deep learning	(i) 17 distinct ICD-9 chapter codes, (ii) 2017 distinct ICD-9 codes	(i) Accuracy of 0.580, (ii) Accuracy of 0.409
Xie et al. [47]	Diagnosis description	Deep learning	2833 ICD-9 codes	Sensitivity score of 0.29, Specificity score of 0.33
Singaravelan et al. [48]	Subjective component	Deep learning	1871 ICD-9 codes	Recall score for chapter code, 0.57; Recall score for block, 0.49; Recall score for three-digit code, 0.43; Recall score for full code, 0.45
Zeng et al. [49]	Discharge summary	Deep learning	6984 ICD-9 codes	F1 score of 0.42
Huang et al. [50]	Discharge summary	Deep learning	(i) 10 ICD-9 codes, (ii) 10 blocks	(i) F1 score of 0.69, (ii) F1 score of 0.72
Our study	Clinical notes	Deep learning	1131 ICD-10 codes	Precision of 0.96, Recall of 0.99, F-score of 0.98

**Strength of the study:** There are a few advantages of this study. The first thing to note is that this is the first study to evaluate the performance of a CNN model built on deep learning for ICD-10 code prediction utilizing SOAP notes and medication lists. Second, our CNN model can accurately identify ICD-10 codes based on SOAP notes and prescription lists, which can assist clinicians in correctly identifying ICD-10 codes. Thirdly, our model offers a list of the top 10 diagnoses with probability scores, with the assumption that a

diagnosis with a higher probability score will be more accurate. Doctors can therefore choose a diagnosis from the diagnosis list.

**Limitation of the study:** There are obviously some drawbacks to this study. First, data from a single university hospital was employed in our model. Data from other hospitals might have an impact on the model's performance because their clinical notes may have different writing styles and disease identification patterns. Second, the dataset we used came from five departments using medical records. The dataset is not particularly large. Third, we did not validate our research with data from an outside source. Lastly, we used only one method in our study.

**Future perspective:** Our deep learning-based CNN model behaves as an assistance tool that helps physicians in better decision making. This model could reduce the manual entry of ICD-10 codes. As our findings are promising, we will use more data to make the model more effective. In the future, we will focus on more advanced NLP methods such as BERT (bidirectional encoder representations from transformers) to predict ICD codes.

## 5. Conclusions

In this study, we built a CNN-based deep learning model to predict ICD-10 codes based on data from the SOAP notes and drug lists of clinical notes. The cardiology department's prediction model performed best with a precision of 0.96, recall of 0.99, and F-score of 0.98. Clinically good performance is achieved by the prediction model. Textual data are often complex, of variable lengths, and with nuances in meaning. Our deep learning models can effectively handle such complexities by capturing contextual dependencies and long-range dependencies within the text. They can capture a text's non-structural (SOAP) and structural information (drug list), enabling a more accurate classification. This model recommends utilizing disease lists chronologically to assist physicians in selecting appropriate ICD-10 codes. This CNN model, which is based on deep learning, performs better than models applied in earlier experiments. Our research established that drug lists and SOAP notes have a significant role in predicting ICD-10 codes. With the help of our deep learning model, we can classify diseases appropriately. A quick and accurate ICD-10 coding decision can be made by doctors with the help of this approach.

**Author Contributions:** Conceptualization, J.H.B.M. and M.-C.L.; Methodology, J.H.B.M. and M.-C.L.; Software, J.H.B.M. and M.-C.L.; Validation, J.H.B.M. and M.-C.L.; Formal analysis, J.H.B.M., C.-C.K. and M.-C.L.; Data curation, J.H.B.M.; Writing—original draft, J.H.B.M.; Writing—review & editing, J.H.B.M., C.-Y.Y., H.-C.Y. and M.-C.L.; Supervision, M.-C.L.; Project administration, M.-C.L.; Funding acquisition, M.-C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by grants from the Ministry of Science and Technology, Taiwan (Grant number. 106-2634-F-038-002, 108-2314-B-038-053-MY3), to J.H.B.M., C.-C.K. and M.-C.L., and this work was supported by Taipei Medical University for M.-C.L. This research is partly sponsored by the National Science and Technology Council (NSTC) under grant NSTC 111-2622-8-038-006-IE and 110-2320-B-038-029-MY3, and the Ministry of Education in Taiwan.

**Institutional Review Board Statement:** The study was approved by the Taipei Medical University Joint Institutional Review Board of Human Research in accordance with the Declaration of Helsinki (Ethical Approval Number. IRB No: N201705069).

**Informed Consent Statement:** Patient consent was not required due to the retrospective cohort study of medical records.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to express our gratitude for the assistance provided by the learning hospital project at Wan Fang Hospital and Taipei Medical University.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- World Health Organization. *The International Classification of Diseases*; 10th Revision; World Health Organization: Geneva, Switzerland, 2015; Available online: <https://icd.who.int/browse10/2019/en> (accessed on 3 March 2023).
- Rae, K.; Britt, H.; Orchard, J.; Finch, C. Classifying sports medicine diagnoses: A comparison of the International classification of diseases 10-Australian modification (ICD-10-AM) and the Orchard sports injury classification system (OSICS-8). *Br. J. Sport. Med.* **2005**, *39*, 907–911. [[CrossRef](#)] [[PubMed](#)]
- Subotin, M.; Davis, A. A system for predicting ICD-10-PCS codes from electronic health records. In Proceedings of the BioNLP 2014, Baltimore, MD, USA, 26–27 June 2014; pp. 59–67.
- Nadathur, S.G. Maximising the value of hospital administrative datasets. *Aust. Health Rev.* **2010**, *34*, 216–223. [[CrossRef](#)] [[PubMed](#)]
- Bottle, A.; Aylin, P. Intelligent information: A national system for monitoring clinical performance. *Health Serv. Res.* **2008**, *43*, 10–31. [[CrossRef](#)] [[PubMed](#)]
- Quan, H.; Sundararajan, V.; Halfon, P.; Fong, A.; Burnand, B.; Luthi, J.C.; Saunders, L.D.; Beck, C.A.; Feasby, T.E.; Ghali, W.A. Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. *Med. Care* **2005**, *43*, 1130–1139. [[CrossRef](#)] [[PubMed](#)]
- Banerji, A.; Lai, K.H.; Li, Y.; Saff, R.R.; Camargo, C.A., Jr.; Blumenthal, K.G.; Zhou, L. Natural language processing combined with ICD-9-CM codes as a novel method to study the epidemiology of allergic drug reactions. *J. Allergy Clin. Immunol. Pract.* **2020**, *8*, 1032–1038.e1. [[CrossRef](#)]
- Adams, D.L.; Norman, H.; Burroughs, V.J. Addressing medical coding and billing part ii: A strategy for achieving compliance. a risk management approach for reducing coding and billing errors. *J. Natl. Med. Assoc.* **2002**, *94*, 430.
- Yang, L.; Kenny, E.M.; Ng, T.L.J.; Yang, Y.; Smyth, B.; Dong, R. Generating plausible counterfactual explanations for deep transformers in financial text classification. *arXiv* **2020**, arXiv:2010.12512.
- Melville, P.; Gryc, W.; Lawrence, R.D. Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 1275–1284.
- Slater, L.T.; Karwath, A.; Williams, J.A.; Russell, S.; Makepeace, S.; Carberry, A.; Gkoutos, G.V. Towards similarity-based differential diagnostics for common diseases. *Comput. Biol. Med.* **2021**, *133*, 104360. [[CrossRef](#)]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the NIPS 2012, Lake Tahoe, NV, USA, 3–8 December 2012.
- Zhang, Y.; Chen, R.; Tang, J.; Stewart, W.F.; Sun, J. LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1315–1324. [[CrossRef](#)]
- Wang, S.M.; Chang, Y.H.; Kuo, L.C.; Lai, F.; Chen, Y.N.; Yu, F.Y.; Chen, C.H.; Lee, C.W.; Chung, Y. Using Deep Learning for Automatic Icd-10 Classification from FreeText Data. *Eur. J. Biomed. Inform.* **2020**, *16*, 1–10. [[CrossRef](#)]
- Chen, P.F.; Wang, S.M.; Liao, W.C.; Kuo, L.C.; Chen, K.C.; Lin, Y.C.; Yang, C.Y.; Chiu, C.H.; Chang, S.C.; Lai, F. Automatic ICD-10 Coding and Training System: Deep Neural Network Based on Supervised Learning. *JMIR Med. Inform.* **2021**, *9*, e23230. [[CrossRef](#)]
- Wang, S.M.; Lai, F.; Sung, C.S.; Chen, Y. ICD-10 Auto-coding System Using Deep Learning. In Proceedings of the 10th International Workshop on Computer Science and Engineering (WCSE 2020), Yangon, Myanmar, 26–28 February 2020; pp. 46–51.
- Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural Language Processing (almost) from Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
- Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 23–24 June 2014; Association for Computational Linguistics: Baltimore, MD, USA, 2014; pp. 655–665.
- Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1746–1751.
- Gangavarapu, T.; Jayasimha, A.; Krishnan, G.S.; Kamath, S. Predicting ICD-9 Code Groups with Fuzzy Similarity Based Supervised Multi-label Classification of Unstructured Clinical Nursing Notes. *Knowl.-Based Syst.* **2020**, *190*, 105321. [[CrossRef](#)]
- Chen, M.; Hao, Y.; Hwang, K.; Wang, L.; Wang, L. Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access* **2017**, *5*, 8869–8879. [[CrossRef](#)]
- Moons, E.; Khanna, A.; Akkasi, A.; Moens, M.F. A comparison of deep learning methods for ICD coding of clinical records. *Appl. Sci.* **2020**, *10*, 5262. [[CrossRef](#)]
- Krishnan, G.S.; Kamath, S.S. Evaluating the Quality of Word Representation Models for Unstructured Clinical Text Based ICU Mortality Prediction. In Proceedings of the 20th International Conference on Distributed Computing and Networking, ICDCN'19, Bangalore, India, 4–7 January 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 480–485.
- Mauch, J.T.; Rios-Diaz, A.J.; Kozak, G.M.; Zhitomirsky, A.; Broach, R.B.; Fischer, J.P. How to Develop a Risk Prediction Smartphone App. *Surg. Innov.* **2020**, *28*, 438–448. [[CrossRef](#)]
- Kavuluru, R.; Rios, A.; Lu, Y. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. *Artif. Intell. Med.* **2015**, *65*, 155–166. [[CrossRef](#)]

26. Li, H.; Li, X.; Ramanathan, M.; Zhang, A. Prediction and Informative Risk Factor Selection of Bone Diseases. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *12*, 79–91. [\[CrossRef\]](#)
27. Jin, B.; Che, C.; Liu, Z.; Zhang, S.; Yin, X.; Wei, X. Predicting the Risk of Heart Failure With EHR Sequential Data Modeling. *IEEE Access* **2018**, *6*, 9256–9261. [\[CrossRef\]](#)
28. Arnaud, É.; Elbattah, M.; Gignon, M.; Dequen, G. NLP-Based Prediction of Medical Specialties at Hospital Admission Using Triage Notes. In Proceedings of the 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), Victoria, BC, Canada, 9–12 August 2021; pp. 548–553. [\[CrossRef\]](#)
29. Vinod, P.; Safar, S.; Mathew, D.; Venugopal, P.; Joly, L.M.; George, J. Fine-tuning the BERTSUMEXT model for Clinical Report Summarization. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 5–7 June 2020; pp. 1–7.
30. Teng, F.; Ma, Z.; Chen, J.; Xiao, M.; Huang, L. Automatic medical code assignment via deep learning approach for intelligent healthcare. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2506–2515. [\[CrossRef\]](#)
31. Bird, S. NLTK: The natural language toolkit. In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, Sydney, Australia, 17–18 July 2006; pp. 69–72.
32. Rehurek, R.; Sojka, P. *Gensim—Python Framework for Vector Space Modelling*; NLP Centre, Faculty of Informatics, Masaryk University: Brno, Czech Republic, 2011; Volume 3.
33. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
34. Chollet, F. Keras. Internet. GitHub. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 11 April 2023).
35. Hu, S.; Teng, F.; Huang, L.; Yan, J.; Zhang, H. An explainable CNN approach for medical codes prediction from clinical text. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 256. [\[CrossRef\]](#)
36. Suo, Q.; Ma, F.; Yuan, Y.; Huai, M.; Zhong, W.; Zhang, A.; Gao, J. Personalized Disease Prediction using a CNN-based Similarity Learning Method. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 811–816.
37. Cheng, Y.; Wang, F.; Zhang, P.; Hu, J. Risk Prediction with Electronic Health Records: A Deep Learning Approach. In Proceedings of the 2016 SIAM International Conference on Data Mining, Miami, FL, USA, 5–7 May 2016; pp. 432–440.
38. Diao, X.; Huo, Y.; Zhao, S.; Yuan, J.; Cui, M.; Wang, Y.; Lian, X.; Zhao, W. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. *Int. J. Med. Inform.* **2021**, *153*, 104543. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Rashidian, S.; Hajagos, J.; Moffitt, R.A.; Wang, F.; Noel, K.M.; Gupta, R.R.; Tharakan, M.A.; Saltz, J.H.; Saltz, M.M. Deep Learning on Electronic Health Records to Improve Disease Coding Accuracy. *AMIA Jt. Summits Transl. Sci. Proc.* **2019**, *2019*, 620–629. [\[PubMed\]](#)
40. Li, M.; Fei, Z.; Zeng, M.; Wu, F.X.; Li, Y.; Pan, Y.; Wang, J. Automated ICD-9 Coding via A Deep Learning Approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 1193–1202. [\[CrossRef\]](#)
41. Choi, E.; Bahadori, M.T.; Schuetz, A.; Stewart, W.F.; Sun, J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *JMLR Workshop Conf. Proc.* **2016**, *56*, 301–318.
42. Hsu, C.C.; Chang, P.C.; Chang, A. Multi-Label Classification of ICD Coding Using Deep Learning. In Proceedings of the International Symposium on Community-Centric Systems (CcS), Tokyo, Japan, 23–26 September 2020; pp. 1–6.
43. Gangavarapu, T.; Krishnan, G.S.; Kamath, S.; Jeganathan, J. FarSight: Long-Term Disease Prediction Using Unstructured Clinical Nursing Notes. *IEEE Trans. Emerg. Top. Comput.* **2020**, *9*, 1151–1169. [\[CrossRef\]](#)
44. Samonte, M.J.C.; Gerardo, B.D.; Fajardo, A.C.; Medina, R.P. ICD-9 tagging of clinical notes using topical word embedding. In Proceedings of the 2018 International Conference on Internet and e-Business, Taipei, Taiwan, 16–18 May 2018; pp. 118–123.
45. Obeid, J.S.; Dahne, J.; Christensen, S.; Howard, S.; Crawford, T.; Frey, L.J.; Stecker, T.; Bunnell, B.E. Identifying and Predicting intentional self-harm in electronic health record clinical notes: Deep learning approach. *JMIR Med. Inform.* **2020**, *8*, e17784. [\[CrossRef\]](#)
46. Hsu, J.L.; Hsu, T.J.; Hsieh, C.H.; Singaravelan, A. Applying Convolutional Neural Networks to Predict the ICD-9 Codes of Medical Records. *Sensors* **2020**, *20*, 7116. [\[CrossRef\]](#)
47. Xie, P.; Xing, E. A Neural Architecture for Automated ICD Coding. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Melbourne, Australia, 2018; pp. 1066–1076.
48. Singaravelan, A.; Hsieh, C.-H.; Liao, Y.-K.; Hsu, J.L. Predicting ICD-9 Codes Using Self-Report of Patients. *Appl. Sci.* **2021**, *11*, 10046. [\[CrossRef\]](#)
49. Zeng, M.; Li, M.; Fei, Z.; Yu, Y.; Pan, Y.; Wang, J. Automatic ICD-9 coding via deep transfer learning. *Neurocomputing* **2019**, *324*, 43–50. [\[CrossRef\]](#)
50. Huang, J.; Osorio, C.; Sy, L.W. An empirical evaluation of deep learning for ICD-9 code assignment using MIMIC-III clinical notes. *Comput. Methods Programs Biomed.* **2019**, *177*, 141–153. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.