2018 IEEE 31st International Symposium on Computer-Based Medical Systems

An initial investigation of the Charlson comorbidity index regression based on clinical notes

Henrique D. P. dos Santos*, Ana Helena D. P. S. Ulbrich†, Vinícius Woloszyn‡ and Renata Vieira*
*School of Technology at Pontifical Catholic University of Rio Grande do Sul,
Email: henrique.santos.003@acad.pucrs.br, renata.vieira@pucrs.br
†Teaching and Research Group at Conceição Hospital Group, Email: anaulbrich@ghc.com.br
‡Informatics Institute at Federal University of Rio Grande do Sul, Email: vwoloszyn@inf.ufrgs.br

***Abstract*—The Charlson comorbidity index (CCI) is widely used to predict mortality for patients who may have many comorbid conditions. Such index is also used as an indicator of the patients' complexity inside a hospital. In this paper, we evaluate a variety of feature extraction and regression methods to predict the CCI from clinical notes. We used a tertiary hospital dataset with 48 thousand hospitalizations featuring the CCI annotated by physicians. In our experiments, Dense Neural Networks with Word Embeddings proved to be the best regression method, with a mean absolute error of 0.51.**

***Keywords*-charlson comorbidity index; clinical notes; feature extraction; regression methods;**

I. Introduction

Electronic Health Records (EHR) play an important role in hospital environments bringing many benefits in terms of patient safety, effectiveness and efficiency of care, patient satisfaction, and care process [1]. Records of health care practices in hospitals generate a rich and large amount of patient information and an intrinsic relation between symptoms, diseases, drug interaction, and diagnoses that could be used for many purposes [2].

Machine learning techniques could predict existing measures in clinical research to tackle life expectation and readmission rate [3], [4]. In addition, EHR information could be used to identify the patients with higher risk of death and focus attention on them. The Charlson comorbidity index [5] is the most extensively studied way to measure patient risk. This index evaluates 19 medical conditions, each weighted between 1 and 6, based on the relative risk of their association with 1-year mortality.

Comorbidities are coexisting diseases to a disease of interest or an index disease, which may directly affect the prognosis of the disease of interest, or indirectly influence the choice of treatment. Comorbidity affects prognosis, therapy, and outcome, and is associated with decreased health outcomes. The Charlson comorbidity index (CCI) is the most widely used comorbidity index [6], [7].

Usually, when the CCI is evaluated, comorbidities are manually completed by the physician in available CCI calculators or in the hospital system itself, when available. Physicians spend time completing the evaluation, which they could be using for patient care, as this information is usually already available in clinical notes. Our work aims to assist physicians in this task using machine learning regression methods and feature extraction using natural language processing (NLP). This technique could also be applied at hospitals with electronic clinical notes but without CCI specialists.

The main contributions of our work are the following:

- Description of text features that best predict the Charlson comorbidity index in clinical notes;
- Evaluation and comparison of prediction models with the best performance in the CCI regression task.

The rest of this paper is organized as follows: Section II presents previous works on predicting comorbidities through clinical notes. Section III describes the dataset used and the experiment setup, followed by the results in Section IV. In addition, we perform a qualitative analysis of the results in Section V. Finally, in Section VI we summarize our contributions and present further research directions.

II. Background

Comorbidities or International Classification of Diseases (ICD) prediction is a common task that uses text information from Electronic Health Records (EHR). Yousefi et al. [8] modeled multiple comorbidities in diabetes using dynamic Bayesian networks with latent variables. Stacki et al. [9] derived correlations, evidence-based likelihood of comorbidity manifestation in EHR. Some studies focus on finding specific diseases in EHR, such as chronic obstructive pulmonary disease and critical limb ischemia [10], [11].

A common approach to derive the Charlson comorbidity index from clinical notes usually has two phases: first, comorbidities or ICDs are extracted from text, and then the results for each disease are summed using the Charlson scores to reach the CCI. The first work to use this approach [12] extracted the ICD from clinical notes to derive comorbidities from 3,662 notes of pneumonia patients. They were able to find thirteen underreported comorbidities in the administrative data. Also following this approach, Salmasian et al. [13] identified the comorbidities and derived the CCI in 100 admission notes with a 0.74 F1-Score. Both works

2372-9198/18/$31.00 ©2018 IEEE
DOI 10.1109/CBMS.2018.00009

used MedLEE commercial software [14] to extract the ICD from clinical notes.

In another paper, Singh et al. [15] developed an algorithm based on query rules to extract Charlson comorbidities from EHR. Then, they compared two extraction strategies to derive CCI scores: first, with the algorithm and, second, by using ICD-9 codes. This approach was validated in 240 patients. Its sensitivity ranged from 91% to 100%.

Our approach tackles this problem from a different perspective. Instead of using a two-phase strategy, we predict the CCI with just one phase, using regression machine learning techniques. In addition, there is no need for a specific software to extract comorbidities in this approach. Here we use natural language processing to extract information from clinical notes as features for the regression methods.

Representing text in features is a common issue to extract information from clinical notes. Tang et al. [16] used word representation techniques to perform biomedical named entity recognition. The f-measure was improved by only 2% when features such as clustering-based, distributional, and word embedding were individually added to the basic features. Our work uses some of these techniques to extract features from clinical notes, as we describe in the next section.

III. Materials and Methods

We design the experiments to evaluate several automated approaches to predict the Charlson comorbidity index. In the experiments, we extract features from clinical notes to use as input for the regression algorithms. In this Section, we cover the dataset used, the features extracted, and the regression methods evaluated.

A. Data Source

We used a large cohort extracted from the administrative hospitalization database from Hospital Nossa Senhora da Conceição (HNSC). HNSC is part of the Brazilian public healthcare system and provides tertiary care.

Table I
Clinical Note Dataset

Clinical Notes	1,551,907
Hospitalizations	48,907
Mean CCI	2.80
Mean number of Words	6,126

Table I shows the amount of data used in our experiments. The data comprises 1.5 million clinical notes from 48.9 thousand hospitalization records annotated with the Charlson comorbidity index between January 2012 and December 2017. Most patients have a low CCI, with 2.80 as the mean. However, HNSC treats many cancer and AIDS cases with higher CCI.

Ethical approval to use the hospital dataset in this research was granted by the Research Ethics Committee of Conceição Hospital Group under the number 71571717.7.0000.5530.

B. Charlson Comorbidity Index

The Charlson indices used for the machine learning training were obtained from the electronic system available at the hospital, filled by the physicians according to Table II.

Table II
Charlson Comorbidity Index Weights for Diseases

Comorbidity or Age	**Weight**
$<$ 50 years	0
50–59 years	1
60–69 years	2
70–79 years	3
$\geqslant$ 80 years	4
Myocardial Infarction	1
Congestive heart failure	1
Peripheral vascular disease	1
Cerebrovascular Disorders	1
Dementia	1
Chronic pulmonary disease	1
Connective tissue disease	1
Ulcer disease	1
Mild liver disease	1
Diabetes without end organ damage	1
Hemiplegia	2
Moderate or severe renal disease	2
Diabetes with end organ damage	2
Malignant lymphoma	2
Leukemia	2
Any non-metastatic solid tumor	2
Moderate or severe liver disease	3
Acquired Immunodeficiency Syndrome (AIDS)	6
Metastatic solid tumor	6

At HNSC, only patients with more than two days of hospitalization were evaluated with the CCI. The patients' age was used in the weight system when the Charlson comorbidity index was greater than or equal to 1. Patients who were hospitalized only in the emergency or intensive care units were not evaluated. Fig. 1 shows the CCI distribution in the HNSC dataset from zero to 18.

There are several hospitalized patients with no severe comorbidities: the Charlson comorbidity index for 18 thousand of them was zero. The CCI of most patients (30 thousand) stood between 1 and 10. In our cohort study, in only 700 patients the CCI was higher than 10 units.

C. Feature Extraction

In this study, we included four types of natural language processing features: two sets of simple TF-IDF features (Unigram and Multigram), a topic representation (LDA), and a set of semantic representation of words (Word Embeddings).

Unigram: Here, at first we stemmed the words in the corpus using the RSLP Stemmer algorithm [17]. Then, we calculated the term frequency-inverse document frequency (TF-IDF) of each unigram. At the TF-IDF training step, stop

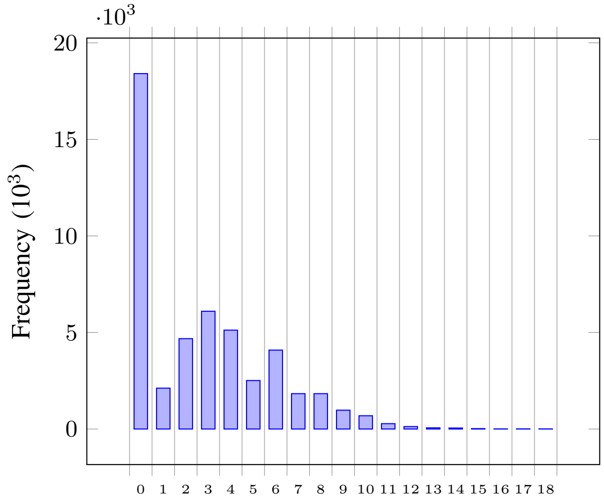

Figure 1. CCI Distribution at HNSC, where the x axis corresponds to the total amount of patients and the y axis represents the Charlson comorbidity index

words were removed, and each word value was normalized with L1 normalization. We limited this features to the 5,000 most representative unigrams.

Multigram: In this approach, we performed the same steps described above but also added bigrams and trigrams with the 5,000 most representative grams.

Latent Dirichlet Allocation: LDA [18] is a generative statistical model that relates each word in a document to a topic, where a topic is a distribution over the set of distinct vocabularies that were found in all documents. Therefore, the content of a topic can be interpreted by verifying the highest probability words in the vocabularies corresponding to the topic. As each word in a document is assigned to a topic, each document could be viewed as a mixture of many topics used to generate that document. LDA was processed with 100 topics, using 10,000 TF-IDF multigram features for these experiments.

Word Embeddings: Word vectors are a way of mapping words in a numerical space. A latent syntactic/semantic vector for each word is induced from a large unlabeled corpus. We used 21 million sentences from HNSC medical records with Word2Vec [19]. The model was trained with 50 dimensions per word and 100 minimum word count. This training resulted in 63 thousand word vectors used as a semantic model in the neural network below.

D. Regression Algorithms

There is a variety of methods that could address the problem of predicting a real value to a document based on its text features [20]. The list below describes the methods we chose in our experiments. The choice took into account algorithms that handle high dimensional data and run parallel jobs.

In all cases, we made a good faith effort to maximize the performance of all methods. In our experiments, we used Scikit-Learn machine learning implementations [21] and Keras for deep learning algorithms [22].

Random Forest: This is a meta estimator that fits a number of classifying decision trees to various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control overfitting [23].

K-Neighbors: It implements learning based on the k nearest neighbors of each query point, where k is an integer value specified by the user [24]. Here we set $k = 5$.

Linear Support Vector Regression: Linear SVR implemented using $liblinear$, providing more flexibility in the choice of loss functions and penalties and being better scaled to large numbers of samples [25].

Neural Networks: Deep learning algorithms are extensively used in biomedical language processing tasks [26]. Neural network algorithms are often associated with word vector representation. In our experiments, we evaluated two types of neural networks commonly used for text processing: Dense and Convolutional neural networks. Every network starts with a embedding input layer for n-gram values and word vector representation when using word embedding features. The embedding layers are then concatenated and fed into a neural network. We briefly describe the neural network used below:

- Dense NN: a densely-connected layer with 128 neurons and $relu$ activation, using $adam$ as optimizer and $mean\ squared\ error$ as loss function;
- Convolutional NN: a convolutional layer with 128 neurons and $relu$ activation, using $stochastic\ gradient\ descent$ as optimizer and $mean\ absolute\ error$ as loss function, and 5 as the length of the convolution window;

All networks were trained over 20 epochs and a batch size of 100. The output layer is a single dense neuron with $relu$ activation.

E. Evaluation

For each regression algorithm, we ran a cross validation with ten stratified folds. The folds were made by preserving the proportion of samples for each Charlson comorbidity index. For every iteration, nine folds were used in the training stage and one fold was used for model evaluation. The mean for all validations was used as the algorithm score. We chose the mean absolute error as the main metric to evaluate the quality of the models.

We use a statistical accuracy metrics, Mean Absolute Error (MAE), to report prediction experiments for it is most commonly used and easy to understand. Less error means that, in average, the algorithm predicted index is closer to the expert actual index

F. Baseline

The baseline is a dummy algorithm that uses the mean of the Charlson index for all instances as the value of all predictions. The mean of the Charlson index in our dataset is 2.80, which generates a mean absolute error of 2.42, as presented in Table III.

IV. Results

We ran all methods against n-grams and LDA, but only the Neural Networks allow the use of word embeddings. The machine learning methods have no ability to process word vector as a feature of the instances. Table III shows the overall results of our experiments.

Table III
MEAN ABSOLUTE ERROR FOR EACH METHOD VS FEATURE

	Unigram ↓	MultiGram	LDA	W.Emb. ↓
Dense NN	**1.32**	1.49	2.42	**0.51**
Random Forest	1.41	1.48	2.42	-
K-Neighbors	1.74	2.04	2.66	-
Linear SVR	1.72	2.15	2.40	-
Convolutional NN	2.41	2.43	2.41	1.36
Baseline	2.42	2.42	2.42	2.42

The best deep learning method, Dense Neural Network (DNN) achieved a mean absolute error of 0.51 using word embedding features. The deep learning methods had a poor performance using n-gram features. Besides the result, DNN requires some overhead: word embeddings need a vast amount of text to train the word vector representation, and the training time of DNN is exponential, higher than the machine learning methods.

As the best classical machine learning algorithm was Random Forest (RF), an ensemble of decision trees with a mean absolute error of 1.41 using unigram features. Random Forest is a good alternative for CCI regression when there is less amount of text to train. The LDA topics approach also got worse results in our experiments for all methods.

A. N-Gram Vocabulary Size

An interesting evaluation for n-gram features is to measure the decrease of the CCI error compared to the number of n-gram features. In Fig. 2, we show the performance of Random Forest algorithms related to the types of n-grams (unigram, bigram, trigram, and multigram) and vocabulary size.

Fig. 2 shows the Random Forest performance for each size of the n-gram vocabulary. When using Unigram, the error difference does not change much after 3.5 thousand words. It is important to evaluate this characteristic to avoid feature overhead.

V. Discussion

Results of this study point to the validity and feasibility of the regression method for identifying the Charlson comorbidity index (CCI) in clinical notes. We were able to calculate the CCI with minimal error using natural language processing (NLP) features, without the need for specialized software to process the texts. As a Charlson comorbidity index prediction, this method can potentially allow automated predictions of patients' outcomes including disability, mortality, length of stay, and readmission [27].

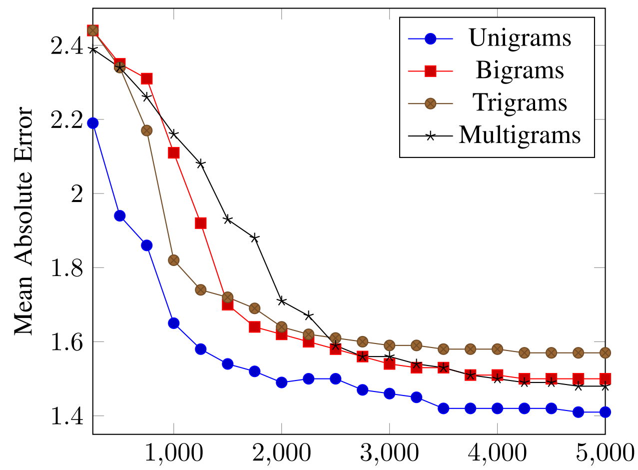

Figure 2. Random Forest Results using N-Gram Features

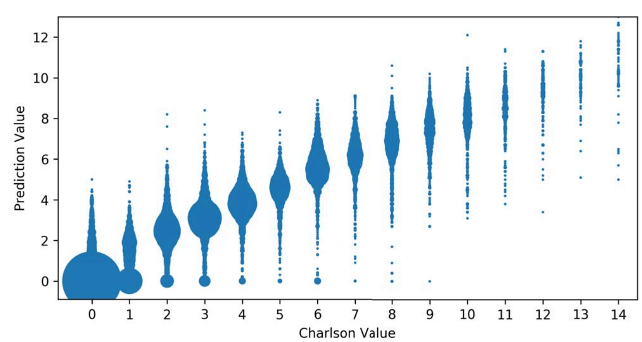

Figure 3. Dense Neural Network Prediction versus CCI Values Defined by Specialists.

Fig. 3 shows the prediction distribution for CCI using Dense Neural Network. Charlson indices higher than 14 were discarded in the figure due to low data density. The chart shows that the error of this method usually stood below the reference standard. The blue circles at zero prediction values and Charlson values 1, 2, 3, 4, and 6 are an effect of the $relu$ activation function in the Dense Neural Network.

Fig. 4 shows the mean absolute errors for each Charlson comorbidity index. The amount of data shown in Fig. 1 is inversely proportional to the mean error. Imbalanced class distribution of a dataset is a serious difficulty for most learning algorithms that assume a relatively balanced distribution [28].

A. Relevant Words

Table IV highlights the most prominent words for the Charlson comorbidity index in clinical notes extracted from

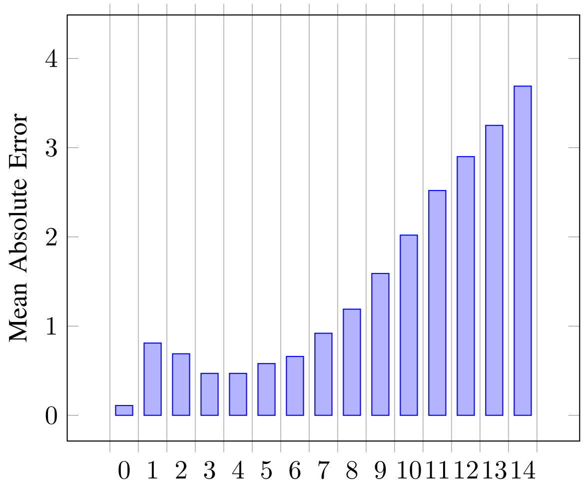

Figure 4. Mean Absolute Error for Each Charlson Comorbidity Index

the Random Forest model. These words are the most heavily weighted words, indicating higher CCI values for these patients.

Table IV
TOP 15 MOST RELEVANT WORDS

Word	Description
Oncology	Department where treat tumor patients
Immunodeficiency	Acquired Immunodeficiency Syndrome (AIDS)
Staging	Metastatic or non-metastatic solid tumor
DM	Diabetes
AIDS	Acquired Immunodeficiency Syndrome (AIDS)
CD4	Immune cell surface, related to AIDS
CRF	Chronic Renal Failure
Cirrhosis	Moderate or severe liver disease
Diabetes	Diabetes
Palliative	Palliative care
SAH	Systemic Arterial Hypertension
Neoplasm	Metastatic or non-metastatic solid tumor
Malignant	Metastatic or non-metastatic solid tumor
Stroke	Cerebrovascular Disorders
COPD	Chronic Obstructive Pulmonary Disease

As expected, the most relevant words are terms related to AIDS (CD4, AIDS, Immunodeficiency) and metastatic tumors (Oncology, Staging, Neoplasm, Malignant). The terms ”DM” and ”Diabetes” show a common disease in hospital patients, which is a risk factor for several other diseases, such as cardiovascular and cerebrovascular diseases. Cardiovascular disease is the main cause of death and disability among patients with diabetes mellitus [29]. Patients in palliative care are patients who have terminal diseases, with poor prognosis. All terms are related to poor prognosis, representing an increase in the Charlson comorbidity index.

VI. CONCLUSION

We were able to derive Charlson comorbidity index automatically from clinical notes using regression methods and textual features with minimal error. This approach could be replicated at other hospitals with the same type of labeled dataset. It is language independent, with no need for licensed software. The Dense Neural Network with Word Embedding outperforms the other methods, but Random Forest with Unigrams could also be a suitable alternative in datasets with lower clinical note density (no enough text to train word vectors). The Portuguese pre-trained word vectors for clinical notes and all experiments are reported at the project’s GitHub page[1] for more details. The main advantages of our method are: assisting healthcare professionals in CCI assignment and using an automated method to derive CCI at hospitals and clinics with electronic health records but no CCI specialists.

Further work could investigate other feature extraction methods as well as other neural network configurations that improve CCI derivation. Character-level and sentence-level models had good performance in Convolutional and Recurrent neural networks for natural language processing. Another experiment should concern on the independence or dependence of the proposed model from the underlying dataset. Also, the same approach could be evaluated to extract ICD and SNOMED codes from clinical notes.

ACKNOWLEDGMENTS

This work was partially supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) Foundation (Brazil), PUCRS (Pontifical Catholic University of Rio Grande do Sul), UFRGS (Federal University of Rio Grande do Sul), and GEP/GHC (Teaching and Research Group at Conceição Hospital Group).

REFERENCES

[1] M. B. Buntin, M. F. Burke, M. C. Hoaglin, and D. Blumenthal, “The benefits of health information technology: a review of the recent literature shows predominantly positive results,” *Health affairs*, vol. 30, no. 3, pp. 464–471, 2011.

[2] P. B. Jensen, L. J. Jensen, and S. Brunak, “Mining electronic health records: towards better research applications and clinical care,” *Nature reviews. Genetics*, vol. 13, no. 6, p. 395, 2012.

[3] A. Hosseinzadeh, M. T. Izadi, A. Verma, D. Precup, and D. L. Buckeridge, “Assessing the predictability of hospital readmission using machine learning.” in *IAAI*, 2013.

[4] J. S. Mathias, A. Agrawal, J. Feinglass, A. J. Cooper, D. W. Baker, and A. Choudhary, “Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data,” *Journal of the American Medical Informatics Association*, vol. 20, no. e1, pp. e118–e124, 2013.

[5] M. Charlson, T. P. Szatrowski, J. Peterson, and J. Gold, “Validation of a combined comorbidity index,” *Journal of clinical epidemiology*, vol. 47, no. 11, pp. 1245–1251, 1994.

[1] https://github.com/nlp-pucrs/cci-regression

[6] Y.-q. Huang, R. Gou, Y.-s. Diao, Q.-h. Yin, W.-x. Fan, Y.-p. Liang, Y. Chen, M. Wu, L. Zang, L. Li *et al.*, “Charlson comorbidity index helps predict the risk of mortality for patients with type 2 diabetic nephropathy,” *Journal of Zhejiang University Science B*, vol. 15, no. 1, pp. 58–66, 2014.

[7] S. R. Austin, Y.-N. Wong, R. G. Uzzo, J. R. Beck, and B. L. Egleston, “Why summary comorbidity measures such as the charlson comorbidity index and elixhauser score work,” *Medical care*, vol. 53, no. 9, p. e65, 2015.

[8] L. Yousefi, L. Saachi, R. Bellazzi, and L. C. A. Tucker, “Predicting comorbidities using resampling and dynamic bayesian networks with latent variables,” in *Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on*. IEEE, 2017, pp. 205–206.

[9] S. B. Stakic and S. Tasic, “Secondary use of ehr data for correlated comorbidity prevalence estimate,” in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 3907–3910.

[10] B. E. Himes, Y. Dai, I. S. Kohane, S. T. Weiss, and M. F. Ramoni, “Prediction of chronic obstructive pulmonary disease (copd) in asthma patients using electronic medical records,” *Journal of the American Medical Informatics Association*, vol. 16, no. 3, pp. 371–379, 2009.

[11] N. Afzal, V. P. Mallipeddi, S. Sohn, H. Liu, R. Chaudhry, C. G. Scott, I. J. Kullo, and A. M. Arruda-Olson, “Natural language processing of clinical notes for identification of critical limb ischemia,” *International Journal of Medical Informatics*, 2017.

[12] J.-H. Chuang, C. Friedman, and G. Hripcsak, “A comparison of the charlson comorbidities derived from medical language processing and administrative data.” in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2002, p. 160.

[13] H. Salmasian, D. E. Freedberg, and C. Friedman, “Deriving comorbidities from medical records using natural language processing,” *Journal of the American Medical Informatics Association*, vol. 20, no. e2, pp. e239–e242, 2013.

[14] C. Friedman, L. Shagina, Y. Lussier, and G. Hripcsak, “Automated encoding of clinical documents based on natural language processing,” *Journal of the American Medical Informatics Association*, vol. 11, no. 5, pp. 392–402, 2004.

[15] B. Singh, A. Singh, A. Ahmed, G. A. Wilson, B. W. Pickering, V. Herasevich, O. Gajic, and G. Li, “Derivation and validation of automated electronic search strategies to extract charlson comorbidities from electronic medical records,” in *Mayo Clinic Proceedings*, vol. 87, no. 9. Elsevier, 2012, pp. 817–824.

[16] B. Tang, H. Cao, X. Wang, Q. Chen, and H. Xu, “Evaluating word representation features in biomedical named entity recognition tasks,” *BioMed research international*, vol. 2014, 2014.

[17] V. M. Orengo and C. Huyck, “A stemming algorithm for the portuguese language,” in *String Processing and Information Retrieval, 2001. SPIRE 2001. Proceedings. Eighth International Symposium on*. IEEE, 2001, pp. 186–193.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[20] S. Mohammad and F. Bravo-Marquez, “Wassa-2017 shared task on emotion intensity,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017, pp. 34–49.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] F. Chollet *et al.*, “Keras,” https://github.com/keras-team/keras, 2015.

[23] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[25] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.

[26] Z. Jiang, L. Li, D. Huang, and L. Jin, “Training word embeddings for deep learning in biomedical text mining tasks,” in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 625–628.

[27] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey, “Comorbidity measures for use with administrative data,” *Medical care*, vol. 36, no. 1, pp. 8–27, 1998.

[28] Y. Sun, A. K. Wong, and M. S. Kamel, “Classification of imbalanced data: A review,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 04, pp. 687–719, 2009.

[29] C. C. L. Wang, C. N. Hess, W. R. Hiatt, and A. B. Goldfine, “Clinical update: cardiovascular disease in diabetes mellitus: atherosclerotic cardiovascular disease and heart failure in type 2 diabetes mellitus–mechanisms, management, and clinical considerations,” *Circulation*, vol. 133, no. 24, pp. 2459–2502, 2016.