# ICD-9 Tagging of Clinical Notes Using Topical Word Embedding

Mary Jane C. Samonte
Technological Institute of the Philippines
Quezon City,
Philippines 1200,
+639110964
mjcsamonte@yahoo.com

Bobby D. Gerardo
Technological Institute of the Philippines
Quezon City,
Philippines 1200,
+639110964
bobby.gerardo@gmail.com

Arnel C. Fajardo
Technological Institute of the Philippines
Quezon City,
Philippines 1200,
+639110964
acfajardo2000@yahoo.com

Ruji P. Medina
Technological Institute of the Philippines
Quezon City,
Philippines 1200,
+639110964
ruji.medina@tip.edu.ph

## ABSTRACT
Medical records, which contains text, has been dramatically increasing everyday. This means that there is a greater need of analyzing health information in a better way. And this can be done through document classification in natural language applications. In this study, we describe tagging of patient notes with ICD-9 codes through topical word embedding in deep learning called EnHANs. We formulate this paper as a multi-label, multi-class classification problem to categorize the ICD-9 codes of a dataset with 400,000 critical care unit medical records. Knowing accurate diagnosis using ICD-9 codes is a vital information for billing and insurance claims. We demonstrate that through the use of topical word embedding model, we learn to classify patient notes with their corresponding ICD-9 labels moderately well than single-label classification.

## CCS Concepts
**Computing methodologies→Artificial intelligence→Natural language processing→Information extraction**

## Keywords
ICD-9 codes; Topical Word Embedding; Hierarchical Attention Networks; Recurrent Neural Network.

## 1. INTRODUCTION
Medical electronic health records are massively increasing as time goes by. There is a great need for those text documents to automatically categorize for better, faster, cheaper and more convenient annotation for better information classification, storage and retrieval. There are numerous number of single-label document processing models have done and many provided promising output. But it is the multi-label classification approach giving better result than the traditional one [1]. Accurate information retrieval is very important in document classification. Information retrieval plays a big role in insurance claims, medical reports, survival analysis, and causal effect inference and mortality prediction that can improve medical and patient care [2][3].

World Health Organization (WHO) was entrusted with the ICD at its creation in 1948. Today, ICD is the product of progressive revisions to adapt with medical progress in practice for classification of the world's medical conditions in a complete and hierarchical way. ICD-9 labelling is an important process that starts from recording diseases to information storage that are beneficial to fact-based decision-making of all its stakeholders. Some of its usefulness includes assessment of patient health matters in their symptoms and treatments, reimbursement practice, resource allocation asset management trends for observing safety and quality policies in deaths and other administrative purposes [4].

Multiple label classification in ICD-9 is a complex task in nature. Patients can have multiple medical conditions and it follows with multiple diagnosis, and thus resulted to multi-class labelling. ICD-9 labeling is unique for different conditions are dependent due to its hierarchical structure. For example, diabetic patients with diabetes mellitus has an ICD-250. ICD-250 is a code used for hospital bills that indicates a diagnosis on a reimbursement claim. Also, diabetic patients can also have other complications like hypertension which is equal to ICD-401.1 as an additional diagnoses. But cardiac arrest can also be labelled as an ICD-429.2 which is the equivalent billable medical code of a patient. This means that ICD-9 coding is a highly complex label distribution task, that not all patients may have the same diagnosis thought with the same condition, because some ICD-9 labels are mutually exclusive, some conditions are rare and some are common where a medical coder must not simply memorized the label but analyze the whole patient record [5][6].

Since this is a broad area of research, there are various studies in tagging patient records with ICD-9. Text classification has proven


*ICIEB 2018*, April 25–27, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-6375-4/18/04…$15.00
https://doi.org/10.1145/3230348.3230357

its importance through early applications in sentiment analysis, document classification and other rule-based systems whether it is supervised or semi-supervised methods [7][8]. Natural language processing (NLP) is a modern computational science that helps accelerate the annotation of unstructured text documents. This is where the advantage of deep learning can be harnessed to overcome the scalability problem of machine learning involving multiple label classification of textual data [9][10].

More specific modern deep learning approach in document classification is called hierarchical attention networks (HAN). HAN is consists of two parts: a word encoder and attention layer; and sentence a sentence encoder and a sentence-level attention layer which uses bidirectional gated recurrent unit (Bi-GRU) and softmax in its last layer for prediction. But HAN model is only capable of multi-class, single-label document classification. A multi-label classification can be a better method in order to arrive with a more accurate ICD-9 label prediction [9][11].

In this paper, we used topical word embedding to allow each word to have different under various topics. By learning topic embeddings and word embeddings separately, each topic represents a pseudo word by concatenating a word in a specific length [12].

## 1.1 Objective of the Study

In this paper, we developed a model for automatic multi-class labelling of ICD-9 codes of patient notes. We modify HAN model and called it EnHANs by assigning topics for each word in the text corpus, and learn topical word embedding in a hierarchical manner.

## 1.2 Scope of the Study

Enhanced hierarchical attention networks (EnHANs) is the modified approach applied in this study for automatic labelling of patient records. This study used word embedding with topic embedding for multi-class, multi-label classification, with an implementation of sigmoid function for the last layer of the neural network model in recognition accurate ICD-9 codes. We used recurrent neural network (RNN) - on bidirectional long-short term memory (Bi-LSTM) approach.

The dataset used in the experiment is MIMIC3 which is generated by MIT Lab for Computational Physiology. MIMIC3 is an openly available dataset containing tables of 40,000 patients of Beth Israel Deaconess Medical Center intensive care units between 2001 and 2011 [13]. All data manipulation and model training was done in Python 3, Keras and TensorFlow 1.0.0 [14].

# 2. REVIEW OF PREVIOUS STUDIES

Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 is the turning point in the high adoption of electronic medical record (EHR) systems. Compared from 2008, the increase of implementing Basic EHR was over 96% from the Office of the National Coordinator for Health Information Technology (ONC) [15][16][17].

Electronic health records (EHRs) are collection of sequential clinical data of a patient's health treatment received in a hospital. The primary reason of creating an EHR is to provide a quality health service gathered from a sequence of every patient notes containing diagnoses, medications, visits and procedures received. Electronic health record systems can be represented by numerical, categorical, medical codes and time series information. Example of numerical information in patient records are age, weight and height while categorical information examples are gender and marital status. International Statistical Classification of Diseases and Related Health Problems (ICD) and Current Procedural Terminology (CPT) are examples of medical codes while vital signs and other recurring procedures received by a patient refers to time series information. There are various systematic reviews done focused on quality improvement of EHR through predictive analysis [18][19][20]. These studies mentioned above agreed that the primary factors of successful EHR prediction analysis are accuracy and interpretability.

Multi-label classification has the purpose of annotating relevant labels in the context of predicting labels for novel documents. Multi-label prediction is the scenario where there is a large number of classes, like ICD-9 codes, and is increasingly relevant in demand for the use of rapidly rising volume of patient records in a real-world applications. Some published researches called this also as extreme multi-label classification due to the number of documents that either should be ranked or tagged [21][22][23].

Machine learning and data mining frameworks has been developed to present various accurate and efficient method in classifying diseases. A study was conducted to identify 300 patients' samples whether it has or hasn't Type 2 Diabetes Mellitus (T2DM) through their EHR from 2012 -2014. Feature extraction and machine learning classifiers like Naïve-Bayes (NB), k-Nearest Neighbors (k-NN) and Support Vector Machines (SVM) were used and resulted to high identification performances [24]. Another diabetic research related was conducted for a systematic review of the applications of machine learning, data mining techniques and tools. Both supervised and unsupervised techniques were employed. SVM classifiers showed as the most successful algorithm in the experiment they conducted [24].

Deep learning has the promise of surpassing the limitations of machine learning. A study on open domain question information retrieval focused on single label text and document classification with question answering and topic labeling. "Deep Learning for Answer Sentence Selection" is an open domain question information retrieval that was made applicable with an approach that does not need any feature engineering and does not involve linguistic data processing. The study offers an answer sentence selection in a binary classification problem [25]. A study on deep neural networks was conducted in order to investigate the performance of question by testing simple genetic algorithm (GA) on hard deep reinforcement learning (RL) compared to contemporary algorithms [26].

The research group of Google is working in the field of Natural Language Understanding (NLU). One of their published papers performed a recurrent neural network (RNN) experiment to convert a written word to its standard spoken word. They used deep learning for text normalization that is focused on parsing treebank in modeling their neural network. The group used LSTM architecture [27]. Other studies applied two channel LSTM which has the forward and backward neural network hidden layers [28].

Topic models are probabilistic method used for document recognition. Topic models work with polysemous words or words with the same spelling but with different meaning by using topic embedding [29]. Adapting topic word is a technique used for document recognition which aims to identify words that describe the topic of the input document [30].

# 3. METHODS

This study deals with the process of automatic ICD-9 codes multi-label classification of patient notes using deep learning. In order to do this, the architecture of hierarchical attention networks but has been modified by adding topical word embedding. The theoretical framework of this study, enhanced Hierarchical Attention Networks (EnHANs), is showed in figure 1.

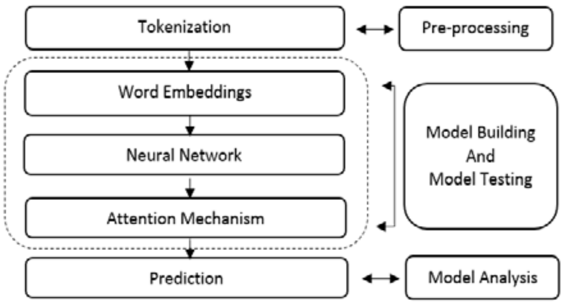


**Figure 1. Theoretical framework of EnHANs.**

## 3.1 Dataset and Tokenization

MIMIC3 (Medical Information Mart for Intensive Care III) is consist of 49,857 patients and the dataset we used in this study. MIMIC3 is all based in ICD-9 codes and it is the actual code corresponding to the diagnosis assigned to the patient for the given row. Aside from the ICD-9 code, MIMIC3 contains unstructured text including medications, patient demographics information, laboratory test results, hourly vital sign measurements, procedures received by the patient, prescribed medications [31][32].

Patient records with the word “Discharge Summary” was processed in the development of the EnHANs model. A requirement of at least one medical diagnosis was used to limit patient records with the same admission ID information and there should have non-empty text. We mapped all codes to their top level representation in the ICD-9, which left 10 top level ICD-9 codes. The data was split into 80-20 training and validation.

In neural networks, deep learning models input are not raw text but vectors or numeric tensors. Vectorizing text is the converting text into numeric tensors. Tokenization is the process of breaking down various units of text, words, or characters, also known as "tokens" [14]. Figure 2 shows a sample of tokenization scheme used in text vectorization from text to tokens to vectors. Word embedding means mapping integer indices or specific words to dense vectors. It can embed sequences of variable lengths but all sequences in a batch must have the same length. Each batch are pack into a single container called “tensor”. Terms that appears at least once in the patient notes are kept in the word-document matrix through Keras Tokenizer [14]. Then vocabulary is selected as the top 10,000 words.

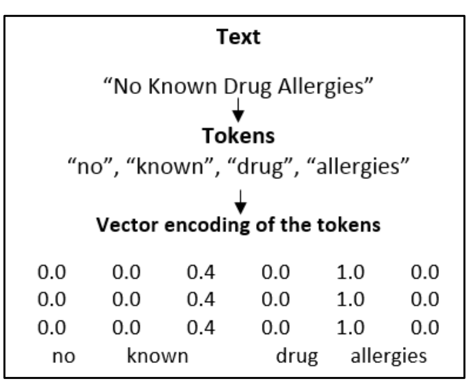


**Figure 2. From text to tokens to vectors.**

## 3.2 Hierarchical Attention Network

The baseline of this study is the Hierarchical Attention Network (HAN) which is made up of a word sequence encoder, a word-level attention layer, a sentence encoder and a sentence-level attention layer that uses softmax function for single-class document classification. Figure 3 shows the HAN model.

## 3.3 Enhanced Hierarchical Attention Network

Our model has the same components of HAN but with topical word embedding and a word input. The architecture was based from multi-input model of question-answering model in Keras [14]. And instead of softmax function, we used sigmoid function for multi-class document classification. Figure 3 shows the EnHANs model of this study.

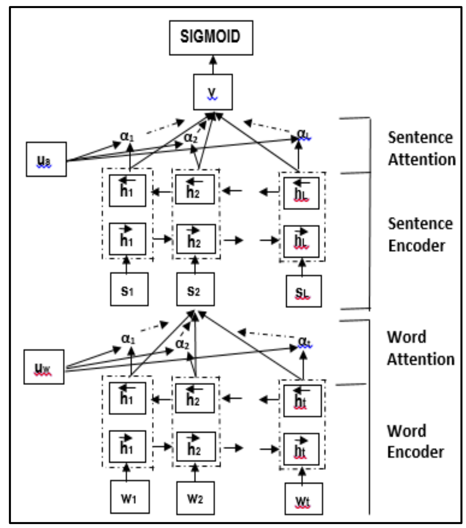


**Figure 3. Hierarchical Attention Network [11]**

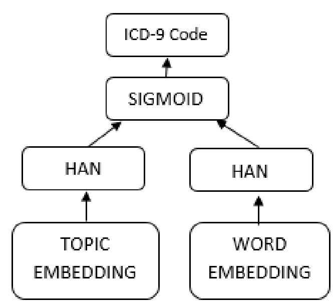


**Figure 4. Enhanced HAN Model.**

## 3.4 Topical Word Embedding

A word represents a certain topic as context and build topical word embedding of (wi; zi) according to topic, wi, and word, zi and used to adapt the average log probability formula indicated in equation 1 [12]. Topical word embedding is done through concatenating wi and zi.

$$\mathcal{L}(D) = \frac{1}{M}\sum_{i=1}^{M}\sum_{-k\leq c\leq k, c\neq 0} \log \Pr(w_{i+c}|w_i) + \log \Pr(w_{i+c}|z_i) \tag{1}$$

## 3.5 RNN in Hierarchical Attention Network

After embedding the text into a sequence of vectors, bidirectional recurrent neural network are used to encode the vectors into a sentence matrix as shown in Figure 5. We keep input vectors separate and normalized at every time step. The sequence of vectors are from first in notes as first vector input. Bidirectional Long Short-Term Memory (Bi-LSTM) computes for the forward hidden sequence xt, the backward hidden sequence kt and the output sequence yt. Through Bi-LSTM, unit yt automatically learn a representation output on both past and future with sensitivity to the input time t. This connects two hidden layers of opposite directions to the same output so that the output layer can get information from past and future states [33].

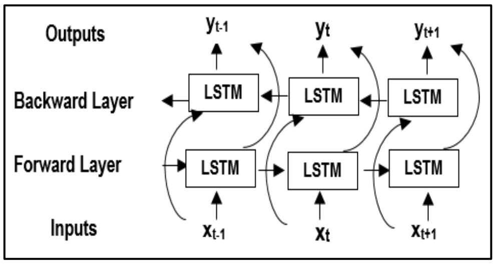


**Figure 5. Bi-LSTM framework in Recurrent Neural Network [33].**

For the 19 label task, we experimented with the number of hidden layers and its size to achieve the best model of 100 sentences per document, 20 words per sentence (each note had a different length of words). We applied sigmoid binary cross entropy as the loss function to improve the model. We used the sigmoid function on the output in predicting labels, where 1 is for all above 0.5 output labels while 0 is for less 0.5. Tanh activation units was used, learning rate for the model is 0.001 and each hidden layer size 100 for the 10 label case and 1000 for the 100 label case.

## 3.6 Model Evaluation

For the validation of results, we use sample based metrics: accuracy, precision, recall and F1 score. Accuracy is the proportion of all true predicted occurrences against all predicted occurrences. An accuracy of 100% means that the predicted occurrences are accurately equal as the actual occurrences. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. F1 score is the weighted average of the precision and recall [6]. Equation 2 shows the formula for accuracy, equation 3 is for precision, equation 4 is for F1 score and equation 5 is for recall formula.

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \tag{2}$$

$$Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Z_i|} \tag{3}$$

$$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2|Y_i \cap Z_i|}{|Y_i| + |Z_i|} \tag{4}$$

$$Recall = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i|} \tag{5}$$

This equation has *Yi* = set of predicted labels, *Zi* = set of ground truth labels, and *n* = number of samples.

# 4. Results and Analysis

In order to overcome the problem with time and hardware availability in this initial study, we deal on tuning the parameters length of 100 sentences per document, 20 words per sentence in the hierarchical attention network (HAN).

The base line in the study is the existing HAN model and we are comparing it to the modified model with topic word embedding, EnHANs. Our setting were a batch size of 64, no dropout, and a learning rate of 0.0001. Table 1 shows the result of the baseline (HANS – no topical word embedding) and EnHANs (with topical word embedding) on 19 label task. Upon running the experiment, the need for regularization was observed in the model, where overfitting occurs, by around the 11th-15th epoch. Compared from the baseline, accuracy was slightly increased in employing topic word embedding, as well as in the case of recall of the model. Accuracy reached 91%, up with 1 point from the baseline model. Unfortunately, precision has maintained strong to the original model. Precision is 10 points higher with the original model. For the F1-score and recall validation, the original and modified model are only the same. The results of recall is both 0.54 while F1-score is both 0.50. This is the result of the performance of the two models that did not undergo pre-processing stage.

**Table 1. Model Performance of the 19 Label Task.**

| | HAN | EnHANS |
|---|---|---|
| Accuracy | 0.900 | 0.900 |
| Precision | 0.760 | 0.910 |
| Recall | 0.540 | 0.540 |
| F1 score | 0.500 | 0.500 |

Upon experimenting with 10 label task, the dataset was underwent pre-processing. this includes removing stopwords and ICD-9 labels, as well as removing words inside brackets [* *] or confidential inofrmation from the patient notes.

Table 2 shows the performance of the baseline model, HAN, in 10 label task with no topical word embedding applied in this experiment.

**Table 2. HAN Performance of the 10 Label Task**

| | HAN |
|---|---|
| Accuracy | 0.830 |
| Precision | 0.765 |
| Recall | 0.626 |
| F1 score | 0.682 |

Table 3 shows the performance of 10 label task with 20 and 6o topics embedded using the formula in equation 1. Based from the result of the experiment done, the number of topics concatenated in each word affects the performance of the ICD-9 codes.

Compared to the result of HAN and the modified model of this study, EnHANs has the potential to greatly outperform the the baseline model, HAN. Through pre-processing that includes removing of stopwords and applying topical word embedding, EnHANs can outperfom the classic hierarchical attention networks.

**Table 3. EnHANs Performance of the 10 Label Task with 20 and 60 Topics.**

| | EnHANs | |
|---|---|---|
| | 20 topics | 60 topics |
| Accuracy | 0.841 | 0.845 |
| Precision | 0.780 | 0.780 |
| Recall | 0.620 | 0.640 |
| F1 score | 0.678 | 0.687 |

# 5. CONCLUSION

There are many applications that explored ICD-9 codes tagging in patient notes. Some are single-label and some are multi-label classification. Deep learning has been known to improve automatic text classification.

MIMIC3 dataset contains big chunks of documents that a lot of processing opportunity is waiting to be done. In this study, we presented the initial result of hierarchical neural network model with no topic embedding, and its performance against the modified base line with topics.

It is important to give emphasis that studies on neural network, especially in deep learning, is resource intensive. The experiment used limited resources in developing the modified framework. Enhanced hierarchical attention network model slightly to moderately outperform the baseline model in some way.

In order to improve this study, further experimental procedures are recommended such as: perform pre-processing procedures in MIMIC3 – this should be done in order to pre-qualify the words in the dataset; explore the number other parameters, that can run in the neural network from time to time; and other researchers can apply additional layer in neural network to observe its performance in labelling ICD-9 codes.

# 6. REFERENCES


[1] L. Lenc, and P. Král, Deep neural networks for czech multi-label document classification, *In Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics*, arXiv preprint arXiv: 1701.03849, 2016.

[2] W. Cleverley, and J. Cleverley, Essentials of health care finance (8th Edition), *Jones and Bartlett Learning*, Burlington, Massachusetts, USA, 2017.

[3] S. Monteith, T. Glenn, J. Geddes, and M. Bauer, Big data are coming to psychiatry: A general introduction, *International Journal of Bipolar Disorders*, 3, 1, 21, 2015.

[4] R. Escorpizo, N. Kostanjsek, C. Kennedy, M. Nicol, G. Stucki, and T. Ustün, Harmonizing WHO's international classification of diseases (ICD) and international classification of functioning, disability and health (ICF): Importance and methods to link disease and functioning, *BMC Public Health*, 13, 1, 742, 2013.

[5] S. Feder, N. Redeker, S. Jeon, D. Schulman-Green, J.A. Womack, J. Tate, R. Bedimo, M. Budoff, A. Butt, K. Crothers, and K. Akgün, Validation of the ICD-9 diagnostic code for palliative care in patients hospitalized with heart failure within the veterans health administration, *American Journal of Hospice and Palliative Medicine*, 1049909117747519, 2017.

[6] Zhang, M., and Zhi-Hua Z. 2014. A review on multi-label learning algorithms, *In Proceedings of the IEEE Transactions on Knowledge and Data Engineering*, 2, 6:8, 1819-1837.

[7] Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H. and Mikolov, T. 2016. Fasttext. zip: Compressing text classification models, arXiv preprint arXiv:1612.03651.

[8] Kavuluru, R., Rios, A., and Lu, Y., 2015. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. *In Journal in Artificial Intelligence in Medicine*, 65, 2, 155-166.

[9] Baumel, T., Nassour-Kassis, J., Elhadad, M., and Elhadad, N., 2017. Multi-label classification of patient notes a case study on ICD code assignment, arXiv preprint arXiv: 1709.09587.

[10] Demner-Fushman, D., and Elhadad, N. 2016. Aspiring to unintended consequences of natural language processing: A review of recent developments in clinical and consumer-generated text processing, *Yearbook of Medical Informatics*, 1, 224.

[11] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. 2016. Hierarchical attention networks for document classification, *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1480-1489.

[12] Y. Liu, Z. Liu, T.S. Chua, and M. Sun, Topical word embeddings, *Proc Twenty-Ninth AAAI Conference on Artificial Intelligence AAAI*, 2418-2424, 2015.

[13] Johnson, A., Pollard, T., Shen, L., Lehman, L., Feng, M., Ghassemi, M., and Mark, R. 2016. MIMIC-III, A freely accessible critical care database, *Scientific Data*, 3, 160035.

[14] Chollet, F. 2017. *Deep Learning with Python*, Manning Publications.

[15] Shickel, B., Tighe, P., Bihorac, A., and Rashidi, P. 2017. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, *In IEEE Journal of Biomedical and Health Informatics*.

[16] Mennemeyer, S., Menachemi, N., Rahurkar, S., and Ford, E., 2016. Impact of the HITECH act on physicians' adoption of electronic health records, *In Journal of the American Medical Informatics Association*, 23, 2, 375-379.

[17] DeSalvo, K., Dinkler, A., and Stevens, L. 2015. The US office of the national coordinator for health information technology: Progress and promise for the future at the 10-year mark, *Annals of Emergency Medicine*, 66, 5, 507-510.

[18] Johnson, A., Altmark, R., Weinstein, M., Pitt, H., Yeo, C., and Cowan, S. 2017. Predicting the risk of postoperative respiratory failure in elective abdominal and vascular operations using the national surgical quality improvement program (NSQIP) participant use data file. *Annals of Surgery*, 266, 6, 968-974.

[19] Kaufman, B., Spivack, B., Stearns, S., Song, P., and O'Brien, E. 2017. Impact of accountable care organizations on utilization, care, and outcomes: A systematic review, *Medical Care Research and Review*, 1077558717745916.

[20] Zheng, K., Gao, J., Ngiam, K., Ooi, B. and Yip, W. 2017. Resolving the bias in electronic medical records, *In Proceedings of the 23rd ACM SIGKDD International*

*Conference on Knowledge Discovery and Data Mining*, 2171-2180.

[21] Babbar, R., and Schölkopf, B., 2017. DiSMEC: Distributed sparse machines for extreme multi-label classification, *In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 721-729.

[22] Feremans, L., Cule, B., Vens, C., and Goethals, B. 2017. *Combining Instance and Feature neighbors for Efficient Multi-label Classification*.

[23] Prabhu, Y., Kag, A., Gopinath, S., Dahiya, K., Harsola, S., Agrawal, R., and Varma, M. 2018. *Extreme Multi-label Learning with Label Features for Warm-start Tagging, Ranking & Recommendation*.

[24] Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., Yang, G., and Chen, Y. 2017. A machine learning-based framework to identify type 2 diabetes through electronic health record, *In International Journal of Medical Informatics*, 1, 97, 120-7.

[25] Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., Yang, G., and Chen, Y. 2017. A machine learning-based framework to identify type 2 diabetes through electronic health record, *In International Journal of Medical Informatics*, 1, 97, 120-7

[26] Yu, L., Hermann, K. M., Blunsom, P., and Pulman, S. 2014. Deep learning for answer sentence selection, *NIPS Deep Learning Workshop*, arXiv preprint arXiv:1412.1632.

[27] Such, F., Madhavan, V., Conti, E., Lehman, J., Stanley, K., and Clune, J. 2017. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning, arXiv preprint arXiv:1712.06567.

[28] Sproat, R., and Jaitly, N. 2016. RNN approaches to text normalization: A challenge, arXiv preprint arXiv:1611.00068.

[29] Rao, K., Peng, F., Sak, H., and Beaufays, F. 2015. Grapheme-to-phoneme conversion using long short-term mermory recurrent neural networks, *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 4225-4229.

[30] Liu, P., Qiu, X., and Huang, X. 2015. Learning context-sensitive word embeddings with neural tensor skip-gram model, *In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI2015)*, 1284-1290.

[31] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. Text summarization techniques: A brief survey, arXiv preprint arXiv: 1707.02268.

[32] Sandfort, V., Johnson, A., Kunz, L., Vargas, J., and Rosing, D. 2018. Prolonged elevated heart rate and 90-Day survival in acutely Ill patients: Data from the MIMIC-III database, *In Journal of Intensive Care Medicine*, 0885066618756828.

[33] Johnson, A., Stone, D., Celi, L., and Pollard, T., 2017. The MIMIC code repository: Enabling reproducibility in critical care research, *In Journal of the American Medical Informatics Association*, 25, 1, 32-39.

[34] Yu, Z., Ramanarayanan, V., Suendermann-Oeft, D., Wang, X., Zechner, K., Chen, L., and Qian, Y. 2015. Using bidirectional LSTM recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech, *In Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding*, ASRU 2015, 338-345.