

## Enhancing the performance of cancer text classification model based on cancer hallmarks

Noha Ali<sup>1</sup>, Ahmed H. AbuEl-Atta<sup>2</sup>, Hala H. Zayed<sup>3</sup>

<sup>1</sup>Department of Computer Science, Modern Academy for Computer Science and Management Technology, Egypt

<sup>1,2,3</sup>Department of Computer Science, Faculty of Computers & Artificial Intelligence, Benha University, Egypt

### Article Info

#### Article history:

Received Oct 25, 2020

Revised Feb 24, 2021

Accepted Mar 27, 2021

#### Keywords:

Biomedical text classification

Cancer hallmarks

CNN

Deep learning

NLP

Phrase embedding

PMCVec

RNN

### ABSTRACT

Deep learning (DL) algorithms achieved state-of-the-art performance in computer vision, speech recognition, and natural language processing (NLP). In this paper, we enhance the convolutional neural network (CNN) algorithm to classify cancer articles according to cancer hallmarks. The model implements a recent word embedding technique in the embedding layer. This technique uses the concept of distributed phrase representation and multi-word phrases embedding. The proposed model enhances the performance of the existing model used for biomedical text classification. The result of the proposed model overcomes the previous model by achieving an F-score equal to 83.87% using an unsupervised technique that trained on PubMed abstracts called PMC vectors (PMCVec) embedding. Also, we made another experiment on the same dataset using the recurrent neural network (RNN) algorithm with two different word embeddings Google news and PMCVec which achieving F-score equal to 74.9% and 76.26%, respectively.

*This is an open access article under the [CC BY-SA](#) license.*



### Corresponding Author:

Noha Ali

Department of Computer Science

Modern Academy for Computer Science and Management Technology

304 Street, Maadi, Cairo, Egypt

Email: cs.noha.ali@gmail.com

## 1. INTRODUCTION

Cancer is a harmful disease that has led to millions of humans deaths. Cancer is regularly depicted within biomedical literature by its hallmarks; A group of related biological behaviors and properties that empower cancer to pass into the body. The major objective of cancer researches is to know the biological tumor mechanisms developments beginning within the body sustained, and turning to be malignant. Six hallmarks of cancer were introduced the first time in the seminal paper published in cell journal [1] then they were extended by another four in this work [2], forming a set of cancer hallmarks that are known till now. The existing set of hallmarks summarizes our knowledge of the disease into a fixed set of changes in cell physiology that influence malignant growth of the tumor (such as evasion of programmed cell death, self-sufficiency in growth signals, sustained angiogenesis, insensitivity to growth-inhibitors, limitless replicative potential and tissue invasion). Over 150k research in cancer published yearly on PubMed. Cancer researchers and oncologists advantage enormously from text mining field information sources in biomedicine such as PubMed. In this paper, we enhance the performance of the classification model [3], which was used to classify PubMed articles based on the 10 hallmarks of cancer.

First, the text classification tasks can accomplish using machine learning (ML) or deep learning (DL) techniques which are both of them under the umbrella of artificial intelligence (AI). DL techniques have the ability to capture the features automatically from the text. On the other hand, ML techniques have to

be fed manually with the extracted features as input. This difference affects the performance of deep learning algorithms making them outperform over ML techniques in the text classification task. Second, the normal (natural) text differs from the biomedical text in the following characteristics, a medical term may be written abbreviated like this cell type name called (OR) means outer root cell type, not that proposition letter. Also, a very important characteristic, the medical term may consist of phrases or compound-words like this protein name (hypoxia-inducible) or symptom like high-blood-pressure all of these characteristics may cause dispersion problems in classification [4].

Despite the achievement of high-quality vector space models, for example, Word2vec and Glove, they just give unigram word representation and the semantics for phrases consist of multi-word must be approximated through the compositional approaches. In biomedical text processing, it is difficult to write technical phrases for symptoms, medications, and diseases as single words to capture the right meaning. To solve this problem, in this work, we use a recently un-supervised technique, that uses the concept of the multi-word (phrase) embedding, called PMC vectors (PMCVec) [5] (which is pertained to biomedical articles) for preprocessing to extract the distributed semantic phrases from cancer's abstracts for better classification performance. The PMCVec was implemented in the embedding layer of the convolutional neural network (CNN) algorithm used for biomedical text classification according to cancer hallmarks. Also, we prove that changing in word embeddings technique can improve the performance of classification and also, compares the convolutional neural networks versus recurrent neural networks on the same dataset using two different concepts on embedding uni-gram embedding and multi-word embedding.

DL algorithms and architectures have already made superior advances in speech recognition, computer vision, and natural language processing (NLP) fields [6]. CNN proposed as the first time for image processing by [7] and still working till now and achieves perfect results in various computer vision tasks such as object detection [8], image classification [9], medical image analysis [10], improving the performance of breast cancer detection [11], and a lot of image processing tasks. Also, CNN was applied to speech recognition, for example, it was used to recognize the baby cry and achieved an accuracy of 78.6% on 5 types of baby cries [12], also, used to recognize speech emotions [13]. However, the convolutional neural network (CNN) is used in general NLP tasks, particularly text classification tasks [14]. There are a huge number of researchers applied the CNN algorithm to detect the polarity of a text, the text may be a sentence, paragraph, or document as well to detect the opinion is positive; negative; or neutral, this step is called sentiment analysis. Also, in this work [15] it's used for sentence-level classification, they applied 4 models of the algorithm on different datasets and the algorithm has improved four of seven tasks which include question classification and sentiment analysis. In the biomedical natural processing (Bio-NLP) topic, this work [16] authors used rule-based features with a knowledge-guided convolutional neural network to classify clinical text. However, A convolutional neural network was applied on clinical notes to categorize text fragments, the system [17] outperformed the other ML approaches by almost 15% while the training dataset contains 4000 sentences and the accuracy was 68%. In [18] authors have achieved 54,79% accuracy while classifying biomedical abstracts published in Ohsumed, and the dataset was contained 11,566 medical abstracts.

Furthermore, in the "Cancer" topic, authors in [19] applied the ML algorithm "support vector machine (SVM)" to classify 1,852 biomedical abstracts according to the ten hallmarks of cancer with manual feature engineering achieving average F-score 69.2% with bag-of-words (BOW) methodology, then, they improve the performance using rich features technique achieving F-score 76.8%. Then, the authors compared the result of SVM with the CNN algorithm in this work [3] and they achieved F-Score 76.6% using Google News word vector. Then, the authors made some modifications in the dataset, filter sizes of the model, and word embedding algorithms which improve their model achieving F-score 81.0% with Chiu-win-2 word vector [20]. The paper is organized as: section 2, describes the proposed method, the experimental setup in this research, and clarifies the dataset used. Section 3 evaluates and discusses the proposed technique. Finally, section 4 shows our conclusion.

## 2. RESEARCH METHOD

### 2.1. Model layers

The proposed model to classify cancer articles based on cancer hallmarks is illustrated in Figure 1. The model consists of CNN algorithm layers, which start from the embedding layer followed by 1 convolution layer, then 1 max-pooling layer, and a dense layer.

The input articles should be pre-processed before entering the CNN layers. In the pre-processing process, we use the PMCVec that extract useful phrases from the text by removing the numbers, then chunk the sentences based on the predefined stop words. Then, filter the phrases initially based on frequency statistics then, rank and filter again the extracted phrases by a ranking algorithm; Information Frequency (Info\_Freq). Then, tagging the phrases by underscores'. After preparing the data in the preprocessing phase

the extracted phrases pass the word embedding layer; the process of mapping the vocabularies into vectors which consist of real numbers using language modeling and feature learning methods in NLP.

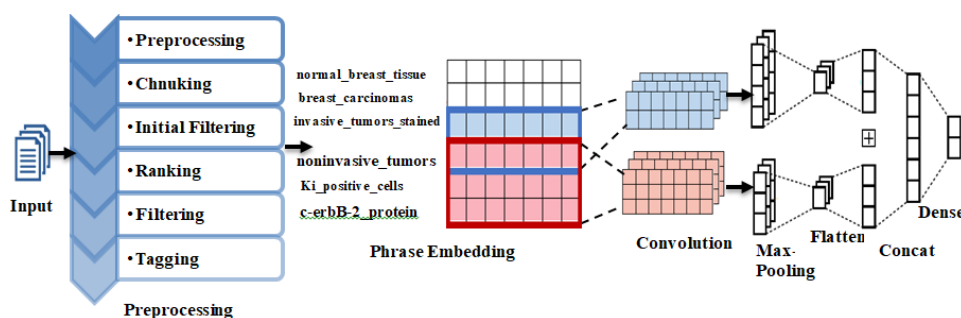


Figure 1. Proposed model

The quality of the word vector can affect the total quality of the text classification. There are a lot of word embeddings available publically like GoogleNews, GloVe and BioNLP. They were mentioned in this survey paper [21] and they compared with PMCvec on five different datasets in [5]. The main differences between them:

- Google News [22]: A popular embedding model used as state-of-the-art, it is trained on Google News dataset. That is a Word2Vec model trained on a general (non-biomedical) corpus. It is a 300-dimensional vector representation.
- GloVe [23]: Combines the power of the Word2Vec model with the effectiveness of the global Co-occurrence statistics method, which is also trained on a general (non-biomedical) corpus of Wikipedia. It is a 300- dimensional vector representation.
- BioNLP [24]: Induced from PubMed, PMC, and their combination using the Word2vec model. It is a 200- dimensional vector representation.
- PMCVec [5]: A Recently word-embedding vectors, which trained on PubMed articles and supports unigram word and multi-word phrases representations. It is a 200- dimensional vector representation.

Therefore, we select the PMCvec because it uses multi-word (phrase) embedding but the other vectors use uni-word embedding. As we mention the biomedical terms, symptoms, and medications are usually written in phrases. So, the PMCvec is the better in our case because the articles in the dataset are about cancer disease which is in the medical domain. After the word-embedding layer, the matrix that contains the values of embedding will enter the convolution layer. Convolutional layer; uses a mathematical model that contains the ReLU activation function (rectified linear unit) that applies the filter sizes to the given text and passes its results to the max-pooling layer in a 2D array. However, in Max-Pooling layer reduces the pooled features to the max by applying a filter matrix. Then, the model should convert the 2-D array to 1-D via the flattening and concat all the 1D-arrays and passes the results to the fully connected layer (dense layer) which is considered as the output layer to decide if the given article positive/negative for the given hallmark. Algorithm 1 describes the steps of the proposed model.

---

**Algorithm 1:** Proposed model for cancer text classification based on cancer hallmarks using CNN algorithm and PMCVec embeddings.

---

- Suppose that  
 $D = \{ \text{label}_1, \text{label}_2, \dots, \text{label}_a \}$  Set of 10 files, one for each hallmark.  
 $Ab_{Tr} = \{ ab_1, ab_2, \dots, ab_n \}$  Set of n abstracts in Training dataset  
 $Ab_{Te} = \{ ab_1, ab_2, \dots, ab_n \}$  Set of n abstracts in Testing dataset

- Training Phase:

- Convert  $D$  into XML format.

**For each file in  $D$**

**For each  $Ab_i$  in  $Ab_{Tr}$**

1. Remove numbers and special characters.
2. Identify noun phrases.
3. Initial filtering by removing any single word occurred once.
4. Ranking using Info\_freq ranking algorithm using this formula for two words phrases:

$$\text{info\_freq}(A,B) = \log \frac{p(A,B)}{p(A)p(B)} * \log(\text{freq}(A,B))$$

, and this formula for 3 word phrases

$$\text{info\_freq}(A,B,C) = \log \frac{p(A,B,C)}{\text{info\_freq}(A,B)p(C)} * \log(\text{freq}(A,B,C))$$

5. Tagging phrases and build embedding matrix.
6. Apply convolution on embedding matrix using different filter sizes.
7. Generate Max-Pooling on each feature map.
8. Flattening (Convert a 2D array into 1D Array).
9. Apply a fully-connected layer with dropout.
10. Save  $Ab_i$  in the trained module.

End For

End For

- Testing Phase:

- For each  $Ab_i$  in  $Ab_{T_e}$  do
  - While EOF ( $Ab_i$ ) do
    1. Load trained module
    2. Evaluate  $S_i$  on the trained module
    3. Calculate the F-score of  $Ab_{T_e}$

End For

## 2.2. Setting model parameters

The proposed model is based on a simple CNN architecture by Kim [15], implementing the neural network (NN) using Keras [25], and TensorFlow was used as a backend tool. The proposed model consists of the PMCvec in the embedding layer followed by one convolution layer of various filter sizes, then 1 max-pooling layer, then finally the output layer. We used the model hyperparameters like the tuned version of S.baker's work [3] except for the embedding layer, where the filter-sizes were 2, 3, 4, number of filters 128, dropout keep probability 0.5, and lambda regularization as default. The training parameters were batch size 64, the number of training epochs 250, and evaluate every 100 steps. Parameters are summarized in Table 1.

Table 1. Model parameters

Parameter	Value
Word Vector Size	200 (Pmcvec)
Filter Sizes	2,3,4
Dropout Probability	0.5
Number of Filters	128
Batch Size	50

## 2.3. Dataset

The same corpus of [19] was used, which contains 1852 biomedical abstracts for training and testing our model. Dataset annotated by an expert with 15+ years of involvement with cancer research. The task is multi-label classification; each abstract may be labeled with zero or more of the ten hallmarks. We split the dataset into 10 Binary-labeled datasets (one for every hallmark), the positive samples in each dataset are the abstracts annotated with that hallmark, where the negative samples are those that aren't annotated with that hallmark. The ten hallmarks are briefly described is:

- Sustaining proliferative signaling: Normal cells need molecules that act as signs for them to grow up and divide. On the other hand, cancer cells, are able to grow up without these external signs.
- Evading growth suppressors: Non-Cancer cells, have operations that can stop the cell growth or division. In Cancer cells, these operations are changed so that they don't deny cell division effectively.
- Resisting cell death: Programmed Cell Death is a technique by which cells can be programmed to die if damaged. But, cancer cells are capable to override these techniques.
- Enabling replicative immortality: Healthy cell dies after a particular number of divisions. But, cancer cells are able to grow and divide endlessly.
- Inducing angiogenesis: Cancer cells; are capable to start angiogenesis, the procedure by which fresh blood vessels are shaped, hence guaranteeing the gracefulness of oxygen and different supplements.
- Activating invasion & metastasis: Cancer-cells can split away from their site of inception to attack encompassing tissue and spread to far off body parts. But, Healthy cells aren't split away.
- Genome instability & mutation: Cancer growth cells for the most part have serious chromosomal variations from the norm, which compound as the illness advances.
- Tumor-promoting inflammation: Aggravation influences the microenvironment encompassing tumors, adding to the multiplication, endurance, and metastasis of malignant cells.

- Deregulating cellular energetics: Cancer cells mostly utilize strange metabolic pathways to create vitality, for example displaying glucose aging in any event, when enough oxygen is available to appropriately breathe.
- Avoiding immune destruction: Non-Cancer cells are visible by the immune system. However, cancer cells aren't.

Furthermore, we change a little bit in the distribution of the dataset than S.baker's work [3]. We divide the annotated data into training, validation, and testing subgroups, 70% for training, 10% for validation, and 20% for testing using a random sampling strategy. Table 2 shows the dataset distribution of positive and negative samples for each hallmark.

Table 2. Dataset distribution

Hallmark	Train		Validation		Test		Total	
	positive	Negative	positive	negative	positive	negative	positive	negative
1 <sup>st</sup>	328	975	43	140	91	275	462	1390
2 <sup>nd</sup>	172	1131	22	161	46	320	240	1612
3 <sup>rd</sup>	303	1000	42	141	84	282	429	1423
4 <sup>th</sup>	81	1222	11	172	23	343	115	1737
5 <sup>th</sup>	99	1204	13	170	31	335	143	1708
6 <sup>th</sup>	208	1095	29	154	54	312	291	1561
7 <sup>th</sup>	227	1076	38	145	68	298	333	1519
8 <sup>th</sup>	169	1143	24	159	47	319	240	1612
9 <sup>th</sup>	74	1229	10	173	21	345	105	1747
10 <sup>th</sup>	77	1226	10	173	21	345	108	1744

### 3. RESULTS AND DISCUSSION

First, we comparing our model using PMCVec embedding versus the CNN model (tuned version) by S.Baker [3]. Figure 2 represents our method outperforms the previous model for each hallmark, and Table 3 compares the F-score percentages for each hallmark individually and the average F-score for both models.

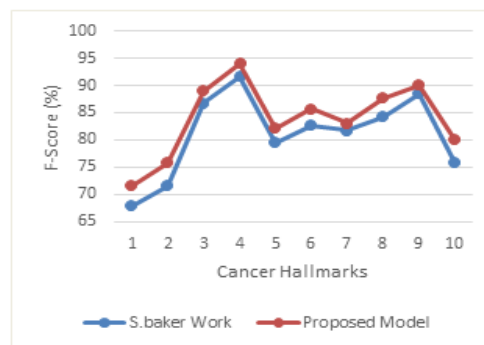


Figure 2. F-score chart comparison for each hallmark using CNN

Table 3. CNN algorithm - test result comparison using the F-score metric

NO.	Hallmark	S.baker [3]	Proposed Model
1	Sustaining Proliferative Signaling	67.90%	71.60%
2	Evading Growth Suppressors	71.50%	75.80%
3	Resisting Cell Death	86.70%	88.90%
4	Enabling Replicative Immortality	91.50%	94%
5	Inducing Angiogenesis	79.40%	82%
6	Activating Invasion & Metastasis	82.60%	85.70%
7	Genome Instability & Mutation	81.70%	83%
8	Tumor-Promoting Inflammation	84.20%	87.70%
9	Deregulating Cellular Energetics	88.30%	90%
10	Avoiding Immune Destruction	75.80%	80%
	Average F-score	81%	83.87%

While evaluating test data in the dataset, the results in Table 3 show that our model outperforms over the existing model for each hallmark individually and on the total average, obtaining an F-score of 83.87% for overall performance which is higher than the previous model that equals to 81%. The proposed model can enhance the existing model by at least 2% to 5% for each hallmark individually and by almost 3% on the total average and if the dataset is larger than the current, the classification results will be better using the multi-word embedding technique. Also, the 4th hallmark result is the highest than the rest, because the examples of the dataset are more relevant to this hallmark than the other. So, the concept of multi-word embedding using PMCvec is effective than the uni-word embeddings techniques. It can improve the performance of embedding, and therefore the result of classification as well.

Another experiment was performed on the same dataset using another DL algorithm; the RNN algorithm to show its performance in text classification tasks in the biomedical domain. Table 4 represents the results of evaluating the test data using the RNN algorithm with two different word embeddings techniques also; uni-gram word embeddings like Google News (default word vector) and phrase embedding like PMCVec.

The result in Table 4 shows that the performance of RNN with PMCVec is 76.26% which is outperforming the RNN with Google News which obtains F-score 74.9%. That is because Google News is trained on general text however the PMCVec is trained on the biomedical text. Also, the phrase embeddings are better than uni-gram embedding. So, PMCvec gives better word embeddings and this is reflected in the classification result as well, but still, both of them are lower than the CNN algorithm result with an average F-score 83.87% as shown in Table 3.

Table 4. RNN algorithm comparison of test result using F-score

No.	Hallmark	RNN (Google News)	RNN (PMCVec)
1	Sustaining Proliferative Signaling	66.0%	68.1%
2	Evading Growth Suppressors	67.4%	69.0%
3	Resisting Cell Death	79.0%	82.1%
4	Enabling Replicative Immortality	82.0%	86.0%
5	Inducing Angiogenesis	74.8%	75.0%
6	Activating Invasion and Metastasis	72.0%	72.4%
7	Genomic Instability and Mutation	76.8%	77.2%
8	Tumor Promoting Inflammation	80.0%	80.3%
9	Cellular Energetics	81.0%	82.0%
10	Avoiding Immune Destruction	70.0%	70.5%
	Average F-score	74.9%	76.26%

Table 5 and Figure 3 compares the benchmarks algorithms using ML algorithms [19], and the DL algorithm [3] with the proposed model using CNN and RNN algorithms on the same dataset. The comparison between them on each hallmark individually and on average of all the hallmarks using the F-score metric.

Based on the previous comparison Table 5 and Figure 3, we conduct that CNN with PMCVec embedding has overcome the other benchmark models in biomedical text classification. The CNN is highly recommended in text classification in the biomedical natural language domain. Also, PMCVec produces higher word embedding performance versus the Google News and Chiu-win-2 with both CNN and RNN algorithms in our case.

Table 5. Comparison between benchmarks algorithms versus the proposed model

Hallmark	ML (SVM + BoW)	ML (SVM+Rich features)	RNN (Google News)	RNN (PMCVec)	CNN (Google News)	CNN (Chiu-win-2)	CNN (PMCVec)
1 <sup>st</sup>	70	67.4	66	68.1	66.3	67.9	71.6
2 <sup>nd</sup>	53.3	65.3	67.4	69	66.7	71.5	75.8
3 <sup>rd</sup>	75.9	82.7	79	82.1	86.9	86.7	88.9
4 <sup>th</sup>	73.1	90.9	82	86	91.2	91.5	94
5 <sup>th</sup>	73.9	85.7	74.8	75	74.8	79.4	82
6 <sup>th</sup>	72.5	72.7	72	72.4	82	82.6	85.7
7 <sup>th</sup>	71.2	69.2	76.8	77.2	72.2	81.7	83
8 <sup>th</sup>	69.9	76.6	80	80.3	81.6	84.2	87.7
9 <sup>th</sup>	78.1	85.7	81	82	76.6	88.3	90
10 <sup>th</sup>	54.3	71.8	70	70.5	67.7	75.8	80
Average	69.22	76.8	74.9	76.26	76.6	81	83.87

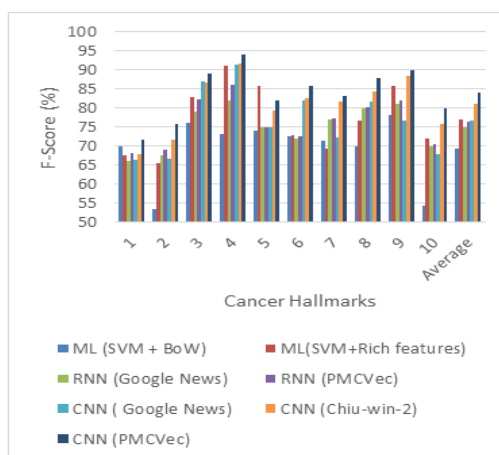


Figure 3. Comparison between benchmarks algorithms

#### 4. CONCLUSION

In this paper, we proposed a model that enhances the performance of the CNN algorithm which is used for text classification of biomedical articles related to cancer disease based on the ten hallmarks of cancer using a new recent concept in the word embedding layer. This technique refers to the use of uni-word and multi-word (phrase) embedding instead of using uni-word embedding only which is suitable for the nature of the medical text. The experimental results show that the concept of the phrase (Multi-word) embedding technique like PMCVec has improved the performance of the existing model achieving an F-score equal to 83.87%, while the previous one was achieved an F-score equal to 81% that uses the uni-word embeddings technique, and if the dataset is larger the classification performance will be better than the current. The proposed model achieving an average F-score greater than other ML and DL models. Also, the results show that CNN is better than RNN in biomedical text classification. Some directions for future work stay open; in addition to changing the word vector, we can examine the effect of changing the optimizer technique, filter sizes, number of filters, or using larger text corpora may offer additional opportunities for enhancement.

#### REFERENCES

- [1] D. Hanahan, R. A. Weinberg, and S. Francisco, "The Hallmarks of Cancer," *cell*, vol. 100, pp. 57-70, 2000. DOI: [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [2] D. Hanahan and R. A. Weinberg, "Review Hallmarks of Cancer: The Next Generation," *Cell*, vol. 144, no. 5, pp. 646-674, 2011. DOI: <https://doi.org/10.1016/j.cell.2011.02.013>
- [3] S. Baker, A. Korhonen, and S. Pyysalo, "Cancer Hallmark Text Classification Using Convolutional Neural Networks," in *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016)*, pp. 1-9, 2016. DOI: <https://doi.org/10.17863/CAM.12420>.
- [4] N. Ali, E. Amer, and H. Zayed, "Understanding Medical Text Related to Breast Cancer: A Review," in *In International Conference on Advanced Intelligent Systems and Informatics*, pp. 280-288, 2017. DOI: 10.1007/978-3-319-64861-3\_26.
- [5] Z. Gero and J. Ho, "PMCVec: Distributed phrase representation for biomedical text processing," *Journal of Biomedical Informatics: X*, vol. 3, p. 100047, 2019. DOI: <https://doi.org/10.1016/j.yjbix.2019.100047>.
- [6] D. Y. Deng, Li, "Deep Learning: Methods and Applications," *Foundations and trends in signal processing*, vol. 7, no. 3-4, pp. 197-387, 2014. DOI: <https://doi.org/10.1561/20000000039>.
- [7] Y. Bengio and Y. Lecun, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks* 3361, vol. 10, Nov. 1995. Available: <https://www.researchgate.net/publication/2453996>.
- [8] L. Jiao, F. Zhang, F. Liu, S. Member, S. Yang, and S. Member, "A Survey of Deep Learning-based Object Detection," *IEEE Access*, vol. 7, pp. 128837-128868, 2017. DOI: 10.1109/ACCESS.2019.2939201.
- [9] H. Education, "Construction of Deep Convolutional Neural Networks For Medical Image Classification," *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 9, no. 2, pp. 1-15, 2019. DOI: 10.4018/IJCVIP.2019040101.
- [10] C. I. S. Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60-88, 2017. DOI: 10.1016/j.media.2017.07.005.

- [11] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, and R. McBride, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific reports*, vol. 9, no. 1, pp. 1-12, 2019. DOI: <https://doi.org/10.1038/s41598-019-48995-4>.
- [12] B. F. Yong, H. N. Ting, and K. H. Ng, "Baby Cry Recognition Using Deep Neural Networks," *World Congress on Medical Physics and Biomedical Engineering 2018*, pp. 809-813, 2019. DOI: [https://doi.org/10.1007/978-981-10-9023-3\\_147](https://doi.org/10.1007/978-981-10-9023-3_147).
- [13] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomedical Signal Processing and Control*, vol. 47, pp. 312-323, 2019. DOI: <https://doi.org/10.1016/j.bspc.2018.08.035>.
- [14] B. C. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification," arXiv preprint arXiv:1510.03820., 2015.
- [15] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, pp. 1746-1751, 2014.
- [16] L. Yao, C. Mao, and Y. Luo, "Clinical Text Classification with Rule-based Features and Knowledge-guided Convolutional Neural Networks," *BMC medical informatics and decision making*, vol. 19, no. 3, arXiv:1807.07425v2, p. 71, 2008.
- [17] I. Li, "Medical Text Classification using Convolutional Neural Networks," *Stud Health Technol Inform* 235, arXiv:1704.06841, pp. 246-50, 2017.
- [18] R. Dollah, C. Y. Sheng, N. Zakaria, M. S. Othman, and A. W. Rasib, "Deep Learning Classification of Biomedical Text using Convolutional Neural Network," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 8, pp. 512-517, 2019. DOI: 10.14569/IJACSA.2019.0100867.
- [19] S. Baker et al., "Automatic semantic classification of scientific literature according to the hallmarks of cancer," *Bioinformatics*, vol. 32, no. 3, pp. 432-440, 2016. DOI: 10.1093/bioinformatics/btv585.
- [20] S. P. Chiu, Billy, Gamal Crichton, Anna Korhonen, "How to Train Good Word Embeddings for Biomedical NLP," in *Proceedings of the 15th workshop on biomedical natural language processing*, pp. 166-174, 2016. DOI: 10.18653/v1/W16-2922.
- [21] S. S. Kalyan, Katikapalli Subramanyam, "SECNLP: A survey of embeddings in clinical natural language processing," *Journal of biomedical informatics*, vol. 101, p. 103323, 2020. DOI: <https://doi.org/10.1016/j.jbi.2019.103323>.
- [22] "word2vec: Tool for computing continuous distributed representations of words." [Online]. Available: <https://code.google.com/archive/p/word2vec/>.
- [23] "Glove: Global Vectors for word representation.", DOI:10.3115/v1/D14-1162, [Online]. Available: <https://nlp.stanford.edu/projects/glove/>.
- [24] "Biomedical Natural Language Processing (BioNLP): Tools and resources." [Online]. Available: <http://bio.nplab.org/>.
- [25] "Keras., Francois Chollet. 2015." [Online]. Available: <https://github.com/fchollet/keras>.