

EFFECTIVENESS OF DEEP LEARNING APPROACH FOR TEXT CLASSIFICATION IN ADAPTIVE LEARNING

^aUmi Laili Yuhana, ^bImamah, ^cChastine Fatichah, ^dBagus Jati Santoso

^{a, c, d}Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

^bDepartment of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember. Surabaya, Indonesia
E-mail: yuhana@if.its.ac.id

Abstract

Nowadays, text classification is extremely beneficial yet required in a multitude of sectors. One area in which text classification is essential for performing the automatic classification of material learning is adaptive learning. Since it is possible for adaptive learning to provide the recommendation over the learning material based on the needs of the student, especially from the aspect of the level of difficulty of the problem. Performing task classification by their difficulty level is done to guarantee that users with low ability levels have access to simple learning resources, and these recommendations will change as the user's abilities grow. Manual text classification always necessitates expert knowledge, whereas adaptive learning necessitates automatic text categorization throughout a set of learning modules. In this study, the performance of Deep Learning in automatically identifying the text of learning materials will be compared. Because the number of terms or words on a single topic of learning material might range from 900 to 1500, Deep Learning was chosen. An IPA subject for junior high school level was gathered from the Teacher's Room and many E-learning platforms in Indonesia for this study. The first step in the classification procedure is labeling the dataset by teacher of biology. In an 80:20 ratio, the dataset was separated into training and testing data. CNN, RNN, and HAN with Word2Vec word embedding were employed in this work. Based on the study, it is known that CNN give the best performance among the three methods with accuracy on training data is 96% and accuracy in validation data is 86%.

Key words: Text classification, adaptive learning, HAN, RNN, CNN, Word2Vec.

INTRODUCTION

Performing classification over the text is highly beneficial and essential in a multitude of sectors. One area that necessitates text classification is adaptive learning. It is a platform of e-learning that tailors the content of the learning to fit the individual needs of its users[1]. The text classification is required in adaptive learning to distinguish the learning material according to its difficulty level.

Adaptive learning research is fascinating, especially in pandemic situations, where the learning process is carried out remotely by an e-learning platform. E-Learning platform has a disadvantage where the learning material is delivered linearly, requiring students to follow a set of resources. The learning recommendation that is based on the students' skills, is not delivered on it. Adaptive learning solves this problem by delivering the learning material according to the level of difficulty to

guarantee that users with inadequate ability levels may receive easy learning materials, and so on. Therefore, Adaptive learning requires the learning material classification stage prior to the recommendation phase.

Text classification is widely researched using various methods. However, different datasets produce different accuracy. Therefore, this study will conduct accuracy tests to find out the most appropriate methods to be used in the learning material dataset to be applied to adaptive learning.

The study will compare the accuracy of three methods of deep learning, they are the Convolutional Neural Network (CNN), the Hierarchical Attention Network (HAN), and the Recurrent Neural Network (RNN). The selection of these three methods is based on the number of terms on a single row of data that reaches 900-1500 words. It is hoped that deep learning will provide good accuracy in the dataset of learning material.

Classification research for education using machine learning and deep learning is fascinating and has received much attention. Text classification as sentiment analysis is widely used to determine student emotion in the Massive Online Open Course[2][3]. Teachers could use this information to provide appropriate treatment to boost student interest and learning engagement[4]. There are also many other classification-related studies, including learner classification based on learning style[5] and student performance classification[6]. Table 1 shows some study on the subject of education that has been done on the classification.

Some of these studies show how artificial intelligence technology can be used to improve student learning outcomes. Adaptive or individualized learning, on the other hand, is a fascinating technique. Table 2 summarizes some adaptive learning research.

In general, adaptive learning research, as shown in Table 2, tries to develop learning suggestions by using multiple ways to create individualized learning paths (personalized learning paths). The purpose is the similar in this study, which is utilizing the complexity of the subject using a text classification approach to generate the learning recommendation for student.

Table 1. Classification research in education fields

Article	Problem of classification	Method	Result
[2]	Educational opinions Classification	CNN, SVC, EvoMSA	Accuracy: CNN=91%, SVC=90%, EvoMSA=93%
[3]	Sentiment Classification to predict learner MOOC satisfaction	Gradient Boosting Trees	F1 Score: 83,75%
[4]	Sentiment classification	SVM	ROC_AUC: 0.90
[5]	Classification to predict learning style	SVM	Accuracy: 85.22%
[6]	Classification of Student performance	Random Wheel	Accuracy: 82%

Table 2. Research on adaptive learning

Article	Purpose of Adaptive Learning	Method	Result
[7]	Improve the excellence of learning by producing adaptive PLP	MapReduce -Based GA	improvement: 107.78%
[8]	alleviate the problem of cold-start in an adaptive learning environment	The combined Random Forest (with Item Response Theory)	AUPR: 0.7952
[9]	User dynamic interests-based adaptive course recommendation	Dynamic Attention Reinforcement Learning	Accuracy: 94.06%
[10]	intelligent adaptive e-learning environment model by embedding the aspect emotional states and the responses of the learner	CNN and Fuzzy Logic.	Median : 0.3106

The utilization over text classification to adaptive learning is proposed in this work. Text classification techniques that are different from text classification in general are required for adaptive learning. The technique of text classification according to difficulty level was recommended in this study to make it easier for the system to deliver learning recommendations based on the user's capacity.

MATERIAL AND METHODS

In this research, we recommend a research scheme as in Figure 1.

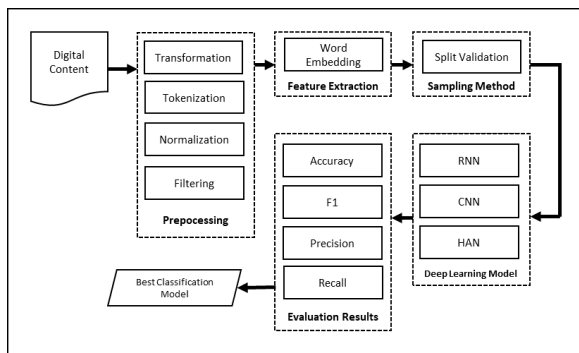


Fig 1. System design of research

The content of biology textbooks gathered via crawling results is the text that this chart uses as input. The structured data is obtained first from the conversion of unstructured data. Additionally, the pegged text will be transformed to matrices by performing two word-embedding techniques: the word2vec and glove. Further comparing the accuracy of the output of each product word embedding method's is the main goal. To produce output, the outcome of word2vec or glove will be processed over the stages of convolution, pooling, and dropout, resulting in class predictions from text input.

Word Embedding

In text processing, word embedding is commonly employed. Its goal is to turn every word into some vectors, with the geometry among these vectors capturing the semantic relationships over words [11]. In Natural Language Processing, word embedding is frequently used for semantic similarity, synonym recognition, concept classification, selection preferences, and analogies. Word2vec and Glove are two well-known word embedding approaches. Word2Vec is the word embedding utilized in this study for text

classification. When processing the text, Word2Vec employs a Neural Network. The input, hidden, and output layers add up to up the process of creating vectors in word2vec. A model trained to predict a given word centered on the context of neighboring words is called Word2vec, it is a modeling that utilizes continuous bag of words (CBOW) and skip-gram scheme. When receiving the input phrases, skip-gram predicts words in terms of timeframes.

Convolutional Neural Network Model

It is a deep neural network that utilizes convolutional filters and further encapsulate features by employing both bias and weight [12][13]. Each layer's output is employed as the input for the next layer. CNN employed one-dimensional classification to classify Science material by difficulty level in this investigation. The one-dimensional CNN is frequently employed to address the problem in the subject of text categorization. Figure 2 depicts CNN's Architecture.

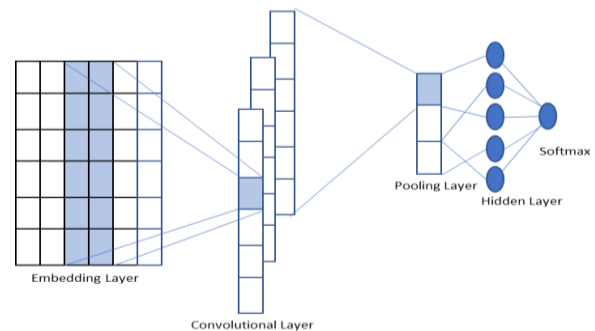


Fig 2. CNN architecture scheme

Embedding layer becomes the first layer of CNN, and it grips the matrix of the sentence of input, which is characterized in the two-dimension matrix: the number of words and the word vector. The embedding layer outputs to the convolutional layer, which has a filter and a kernel that moves vertically along the whole input matrix. This task is done by utilizing the function of activation to build the feature map. Pooling layer then created by employing the greatest value over the feature map, which is then molded into a vector and connected to the output layer to be categorized.

Recurrent Neural Network Model

There exist a form of Artificial Neural Networks (ANN) architecture specifically designed to process sequential data, which

called Recurrent Neural Networks (RNN) [14]. The one that distinguishes RNN from ordinary ANN is the fact that Recurrent Neural Networks (RNN) does not just waste the information from the past during the learning process. RNN is able to store memory (feedback loop) which makes it possible to recognize data patterns well, then use them to make accurate predictions. The ability of RNN in storing information from the past is by looping inside its architecture as seen in the figure 3, which automatically keeps information from the past stored. The idea behind this RNN architecture is how to exploit sequential data structures. The name RNN comes from the fact that it operates repeatedly. This means that the similar task is done for each element of a sequence, and its output depending on the current input and the previous operation. The bottom line is that RNN emphasizes on the nature of data, where the preceding one- or current-time instance (t) affects the instance at the next time ($t+1$).

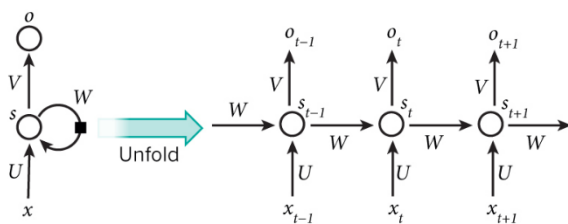


Fig 3. RNN Architecture Model.

Hierarchical Attention Network Model

Hierarchical Attention Network is a deep learning method that is extensively used to perform text classification[15]. HAN used with general architecture which is widely used as in Figure 4 on this study. This architecture only defines hierarchical structures and accommodates different types of encoders and attention.

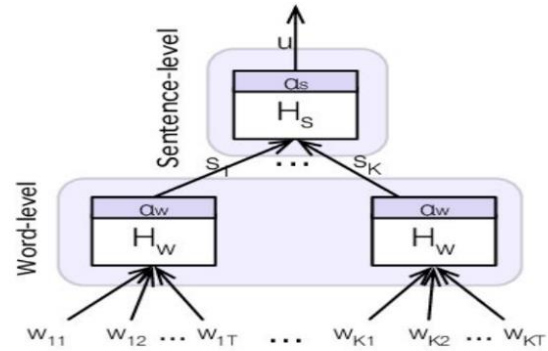


Fig 4. HAN architecture model

RESULT AND DISCUSSION

Dataset

The data was gathered by crawling many Indonesian e-learning sites, one of which being *Ruang Guru* site. Biology is used as a research subject for high school students in grades 1 through 3. The example of dataset as seen in the figure 5. Data is also gathered through printed books that are utilized as reference resources in high school, in addition to crawling on the E-Learning platform. The following is the procedure for gathering datasets:

- The average number of chapters in a book is eleven. Two to five conversation topics are presented in one chapter.
- Material learning is organized by topic and then designated by difficulty level by a Biology's Teacher.
- Crawling data just must be tagged with the level of difficulty because it has been segregated based on the topic of conversation.
- On average there are 400-1000 words more at every material data, so the method used is Deep Learning, which is very appropriate to use. Given the number of features in the data and the benefits of the Deep Learning method, which does not require feature extraction in the learning feature process.

url	materi	topik	subtopik	Tingkat Kesulitan
0 https://blog.ruangguru.com/organi-pemajaan-da	Kamu sadar ngga, kalau kita tuh berpesta raya...	sistem pemajaan manusia	organ pemajaan	Bedang
1 https://blog.ruangguru.com/proso-pemajaan	Artikel ini membahas tentang proso pemajaan...	sistem pemajaan manusia	memasukkan pemajaan	Bedang
2 https://blog.ruangguru.com/kesan-gerak-pada-ma	Squid, tahukah kamu bahwa semua makhluk hidup...	gerak benda dan makhluk hidup	konsep gerak, gerak pada hewan, gerak pada tumbuhan	Bedang
3 https://blog.ruangguru.com/tuang-penyuntan	Coba dan kamu berikan ke cemris, lalu peralat...	gerak benda dan makhluk hidup	alat gerak manusia	Bedang
4 https://blog.ruangguru.com/edel	Artikel Biologi kelas 6 ini membahas segala se...	gerak benda dan makhluk hidup	alat gerak manusia	Bedang
5 https://blog.ruangguru.com/c-penc-peng-an-da	Sewa Siga Corona Artikel Biologi kelas VII...	struktur dan fungsi tumbuhan	jaringan tumbuhan	Mudah
6 https://blog.ruangguru.com/sistem-pemajaan	Artikel Biologi kelas 8 ini akan membahas tentang...	sistem pemajaan manusia	organ pemajaan	Bedang
7 https://blog.ruangguru.com/penggunaan-organ-penc	Artikel ini membahas permasalahan berbagai gangguan...	sistem pencernaan manusia	gangguan pencernaan	Bedang
8 https://blog.ruangguru.com/sistem-pendaran-darah	Artikel ini menjelaskan tentang sistem peredaran...	sistem peredaran darah manusia	komponen sistem peredaran darah, mekanisme per...	Sulit
9 https://blog.ruangguru.com/komposisi-dan-fungs	Squid, kamu tahu gak kalau itu dari total berat...	sistem peredaran darah manusia	komponen sistem peredaran darah	Sulit
10 https://blog.ruangguru.com/sistem-ekresi-manusia	Artikel ini akan membahas tentang fungsi dari s...	sistem ekresi manusia	organ ekresi	Sulit

Fig 5. Learning Material Dataset

The total quantity of data is eighty percent alienated from five hundred to be the training data, and the remaining one is utilized for testing purpose. This information was gathered from a variety of E-Learning sites, including iniitu.id, kelaspintar.id, smpn3konsel.sch.id, RuangGuru, studiobelajar.com, and others.

Text Preprocessing

Cleansing is performed by eliminating special characters: commas, dots, numeric values, and exclamation marks from datasets recorded in CSV format. Datasets that have been free of punctuation marks and numerals will be tagged, allowing each word that makes up sentences to be assigned to a certain topic. The tokenization method produces a set of array data as a result.

Word Embedding

Deep learning for text classification is needed an embedding layer, and it employed word embedding techniques for text categorization. The original idea behind word embedding was that words with semantically similar meanings would frequently appear in the same set of terms. Word embedding is a technique to capture the word context according to geometric properties of the similarity vector. Light and light occupy nearby vector locations because they are semantically comparable. This vector will be utilized as a deep learning input

Evaluation Criteria

Accuracy was used to evaluate the performance of CNN, RNN and HAN. Concurrently, the whole experimental task utilizes the accuracy, with the goal of determining the input percentage of the correct marking of classifier on the test set. Moreover, the test data signifies 20% of the overall data.

Experiment

The study compared the performance of CNN, RNN and HAN. Both of the dimensions used are 100 dimensions. The computation time of the RNN, HAN, and CNN algorithms will be intended in addition to measuring performance based on accuracy. Both tests are used to determine which model is most suited to be utilized in the adaptive learning systems. The shorter the computation period and the greater the accuracy, the better the model for use in adaptive learning environments.

Table 4 shows that CNN approach provides an accuracy loss of 0.69 less than RNN, and 0.09 less than HAN. Validation loss of CNN is 0.51 less than RNN, and 0.28 less than HAN. Its phenomena is indicating that CNN has a better model than RNN and HAN. Training accuracy of RNN and HAN has inferior value of CNN. The accuracy of training and validation of CNN outperformed both approaches, while RNN has a lower value than other methods. Moreover, CNN-Word2vec approach also takes less time to compute than the RNN and HAN.

Table 4. Performance Analysis of CNN, RNN and HAN

Performa nce	CNN- Word2 Vec	RNN- Word2 Vec	HAN- Word2 Vec
Loss acc	0.34	1.03	0.43
Loss Val	0.55	1.06	0.83
Training Acc	0.96	0.51	0.83
Validation Acc	0.86	0.45	0.55
Time	61.20 s	6932.29 s	813,47s

Figures 6, 7, 8, 9, 10, and 11 illustrate the CNN, RNN and HAN method's training and validation graphs.

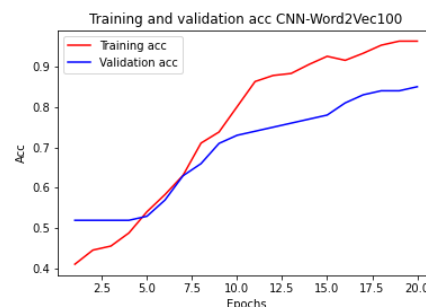


Fig 6. Training and Validation Accuracy of CNN-Word2Vec

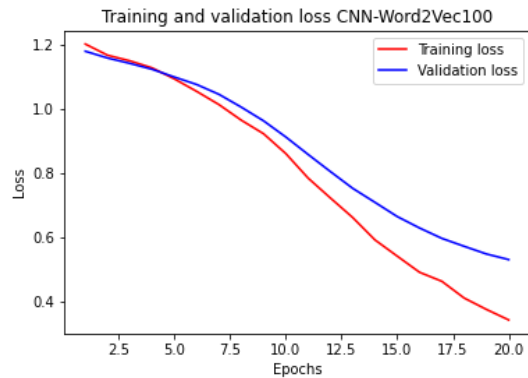


Fig 7. Training and validation Loss of CNN-Word2Vec

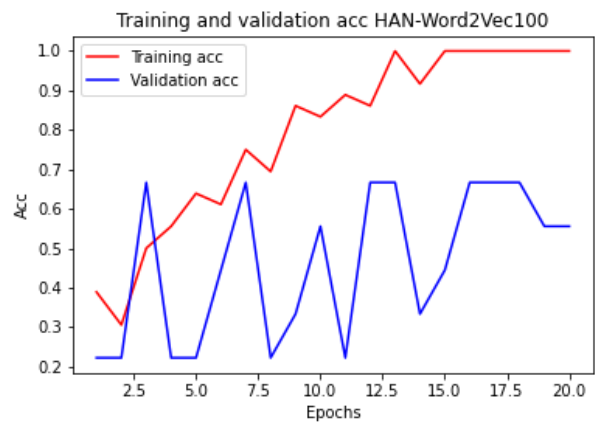


Fig 10. Training and validation Accuracy of RNN-Word2Vec

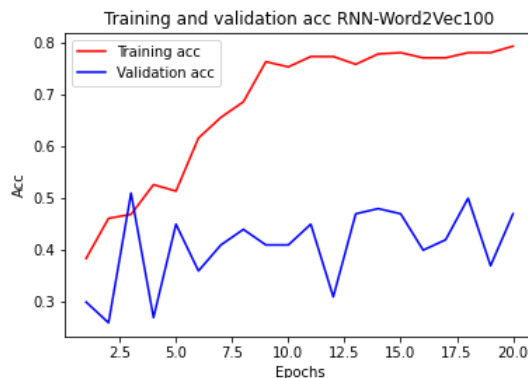


Fig 8. Training and validation Accuracy of RNN-Word2Vec

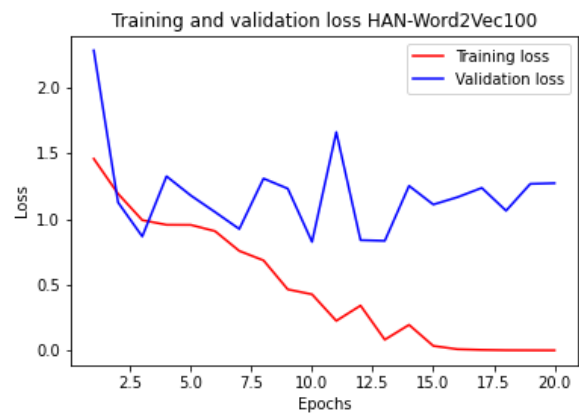


Fig 11. Training and validation Loss of RNN-Word2Vec

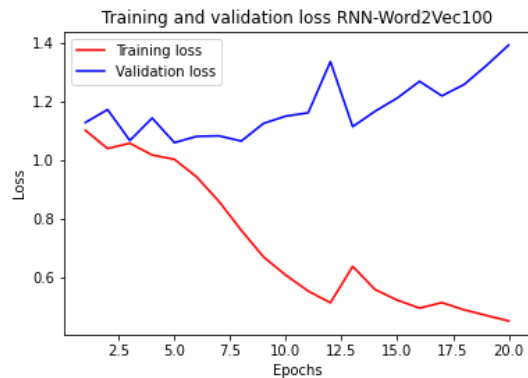


Fig 9. Training and validation Loss of RNN-Word2Vec

CONCLUSION

According to the research, by using word2vec as word embedding be discovered CNN has a higher accuracy than RNN and HNN. CNN has an accuracy of 0.96 or 96 percent. However, the alteration in accuracy and computing time between the three methods is negligible and modest. Its application is very dependent on the dataset and the architecture of

deep learning. Hyperparameter tuning can be used in future research to perceive the influence of parameters on increasing the accuracy.

REFERENCES

- [1] A. Kika, L. Leka, and E. Loka, "Adaptive learning and testing in moodle," *CEUR Workshop Proc.*, vol. 2872, pp. 24–28, 2021.
- [2] M. L. Barrón Estrada, R. Zatarain Cabada, R. Oramas Bustillos, and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Syst. Appl.*, vol. 150, 2020, doi: 10.1016/j.eswa.2020.113265.
- [3] W. Xing, H. Tang, and B. Pei, "Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs," *Internet High. Educ.*, vol. 43, no. 100690, 2019.
- [4] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach.," *Comput. Educ.*, no. 103724, 2019.
- [5] F. Rasheed and A. Wahid, "Learning style detection in E-learning systems using machine learning techniques," *Expert Syst. Appl.*, vol. 174, no. September 2020, p. 114774, 2021, doi: 10.1016/j.eswa.2021.114774.
- [6] A. Khan, S. K. Ghosh, D. Ghosh, and S. Chattopadhyay, "Random wheel: An algorithm for early classification of student performance with confidence," *Eng. Appl. Artif. Intell.*, vol. 102, no. April, p. 104270, 2021, doi: 10.1016/j.engappai.2021.104270.
- [7] M. Birjali, A. Beni-Hssane, and M. Erritali, "A novel adaptive e-learning model based on Big Data by using competence-based knowledge and social learner activities," *Appl. Soft Comput. J.*, vol. 69, pp. 14–32, 2018, doi: 10.1016/j.asoc.2018.04.030.
- [8] K. Pliakos, S. H. Joo, J. Y. Park, F. Cornillie, C. Vens, and W. Van den Noortgate, "Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems," *Comput. Educ.*, vol. 137, no. April, pp. 91–103, 2019, doi: 10.1016/j.compedu.2019.04.009.
- [9] Y. Lin, S. Feng, F. Lin, W. Zeng, Y. Liu, and P. Wu, "Adaptive course recommendation in MOOCs," *Knowledge-Based Syst.*, vol. 224, p. 107085, 2021, doi: 10.1016/j.knosys.2021.107085.
- [10] M. Megahed and A. Mohammed, "Modeling adaptive E-Learning environment using facial expressions and fuzzy logic," *Expert Syst. Appl.*, vol. 157, p. 113460, 2020, doi: 10.1016/j.eswa.2020.113460.
- [11] S. Li, R. Pan, H. Luo, X. Liu, and G. Zhao, "Adaptive cross-contextual word embedding for word polysemy with unsupervised topic modeling," *Knowledge-Based Syst.*, vol. 218, p. 106827, 2021, doi: 10.1016/j.knosys.2021.106827.
- [12] S. Liu and I. Lee, "Sequence encoding incorporated CNN model for Email document sentiment classification," *Appl. Soft Comput.*, vol. 102, p. 107104, 2021, doi: 10.1016/j.asoc.2021.107104.
- [13] A. Khumaidi, E. M. Yuniarno, and M. H. Purnomo, "Welding defect classification based on convolution neural network (CNN) and Gaussian kernel," in *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2017, pp. 261–265.
- [14] W. Q. Yan, *Computational Methods for Deep Learning: Theoretic, Practice and Applications*. 2021.
- [15] N. Pappas and A. Popescu-Belis, "Multilingual Hierarchical Attention Networks for Document

Classification,” September, 2017,
[Online]. Available:

<http://arxiv.org/abs/1707.00896>.