Deep Learning-Based Complaint Classification for Indonesia Telecommunication Company's Call Center

Shinta Devi Lukitasari[1], Fadhil Hidayat[2]
{lukitasari@telkom.co.id[1], fadhil@stei.itb.ac.id[2]}

School of Electrical Engineering and Informatics Bandung Institute of Technology Bandung, Indonesia[1,2]

Abstract. The preliminary research was held to utilize the call center conversations records from a broadband telecommunications company in Indonesia. There is a need from the company to classify customer's complaints automatically by a system to minimize human errors and at once streamline the business processes and resources. Natural Language Processing (NLP), as an integral part of artificial intelligence (AI), empower machines to understand human languages for performing beneficial tasks. The growth of deep learning is the main driver behind NLP for performing various practical applications and business, therefore deep learning is expected to overcome the problems encountered. This paper explains the methods used in designing the classification systems based on deep learning. A literature review is conducted to find the proper algorithm used in classifying problems. Furthermore, it also explains the stages performed in preparing the data and building the system model. From experiments conducted, it can be stated that the RNN algorithm can be used in the classification of customer complaints with the results shown by the accuracy value of the model.

Keywords: Call center, Deep learning, Intent, Complaint, Classification, RNN

1 Introduction

For service providers, a call center is a communication bridge between customers and providers that can handle any inquiries and requests. The development of technology and the way people interact with forces service providers to provide various channels for customers to get in touch with providers, but The CallMiner Index in Q2 2018 [1] data shows that 72% of telecommunication companies' customers prefer the call center to solve their problems than self-service channels, such as chat, email or social media. The survey also showed that 57% of customers tended to think about switching to competitors when they had unpleasant experiences by contacting the call center. The survey results also reflect telecommunication customers in Indonesia. Customers prefer to complain directly to the call center and can easily decide to unsubscribe if they are not satisfied with the call center service. Therefore, to retain customers from the churn, companies need to provide innovation in call center services to deliver excellent customer experience.

Based on the observations of a telecommunication company in Indonesia, several call centers related problems were obtained as follows. First, the company needs to provide a sufficient number of agents and channels so that customers do not need to wait for a long time when they want to contact the call center. Secondly, the running of business processes in handling complaints, among others, depends on the accuracy of the call center agent in determining the classification of problems. The fallacy of the call center agent determines the classification of the customer's complaint can reduce the speed of the company in handling the issues. Last but not least, every call of the call center, incoming and outgoing, will be recorded and stored in a database, and the file should not be deleted for a certain period, in regularly 3 or 5 years. For that reason, an extensive capacity of infrastructure for storage and a significant cost for maintenance are needed. The application of information technology should be utilized to make the data stored are not only cost-based but also can provide more valuable information for the company's business.

Various studies have been performed using artificial intelligence technology to process customer complaint data to increase the company's business [2-9]. Machine learning is an applied branch of artificial intelligence that focuses on developing a system that can replace the capability of humans in the decision making process through the learning process. Deep learning methods have dramatically improved machine learning capabilities, both in terms of speed and accuracy, so they are widely used for Natural Language Processing applications [10]. Natural Language Processing (NLP) is the process of translating the human language to interact with machines. NLP is a method for machines to analyze, understand, and obtain meaning from human language in a smart and useful way so that the machine can perform automated tasks [11]. The application of deep learning in NLP can be used to overcome call center related problems through the classification system.

This paper presents the literature review of a deep learning algorithm for speech and text data classification methods, along with the excellence of each algorithm used on previous research, so it could be applied in further research. Due to the ultimate goal of the entire research is to develop a fully automatic of Telecommunication Company in Indonesia, deep learning algorithm would be an ideal technique, considering its ability to processes data fast and accurate, and also supported by the growth of computing device industry. Thus, the adoption of the system proposed can reduce the risk of errors in the call center, streamline the business processes needed to handle complaints and provide various knowledge for companies to improve service quality.

In research-based on deep learning, the training data preparation is the most important stage besides building a model of the deep learning algorithm itself, therefore this paper will explain the methods used in preparing raw data into training data for classification and building system models. Furthermore, the development of an NLP system based on deep learning will adopt the results of a literature review, initiated with speech recognition and finally results in a classification system. At the end of this paper will present the results of tests conducted on training data and developed classification systems which will be displayed in the system accuracy table.

2 Methodology

For this research, we choose the Design Research Methodology (DRM) [12] as the methodology with considering that the methods provide conscientious stages approach by

planning and implementing design research. The Methodology consists of four stages: Research Clarification (RC), Descriptive Study I (DS-I), Prescriptive Study (PS) and Descriptive Study II (DS-II). In the RC phase, the overall objectives of the study were formulated, and existing knowledge about background problems was explored. At the DS-I stage, empirical data and research literature are collected, and more specific criteria for measuring research objectives are formulated. Using the output, the design is developed and tested at the PS stage. At the DS-II stage, the results can be evaluated according to the evaluation criteria, as well as the applicability of the results.

RC stage has been done through observation of ongoing business process and interviews with the parties involved. The results show that there is a need for technology adoption to make business processes in the call center more efficient. Observations of data samples reinforce validation of the problem. At least 1 of 3 customer conversations stated that the previous complaint had not been appropriately resolved. One of the reasons is the inaccuracy of labeling problems so those customer complaints are not conveyed to the party who should be responsible. This problem is expected to be resolved using deep learning technology to develop customer complaint classifications. The results of preliminary research obtained in the DS-I and PS stages will be explained in the next section.

3 Related Works

Then the literature review was conducted at the DS-I stage to analyze the solutions for a similar problem from previous studies. Various researches have explored the call center records with multiple objectives, including building Automatic Speech Recognition [2, 3] and Speech Act Detection [4], analysis of emotions and customer satisfaction [5, 6], complaint distribution and analysis [7, 8], and others [9]. Table 1 resumes the methods of speech-transcribing, data modeling and classification model that are used in previous studies.

A most common way to perform complaint classification from audio data is to process the output transcripts of speech recognition through a text-based system [13], as seen in several previous studies. Fig. 1, which is redrawn from the summary of various references, gives an overview of how the baseline classification system works. The classification system uses generated speech transcripts which will be processed by the classifier. The speech transcripts could be generated automatically by Automatic Speech Recognition (ASR) tools or manually transcribed. The limitation of using the transcript generated by ASR is that the accuracy of the classifier developed becomes dependent on the performance of ASR tools especially the Word Error Rate (WER) resulted.
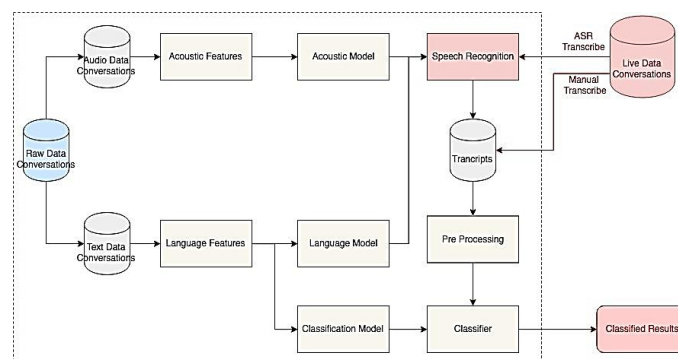
Fig. 1. The baseline of the speech classification system.

Table 1. Methods on previous researches of call center data.

Ref	Features & Model Used	Speech-Transcribing	Classification Method	Objectives
[2]	Linguistic Features: N-gram Model	Manual	-	ASR Model Data Preparation
[3]	Acoustic Features: GMM, TDNN Linguistic Features: N-gram Model	Automatic	Statistical Keyword Score	ASR Development
[4]	Acoustic Features: Pitch, Duration, Intensity, Rhythm	-	Logistic regression	Dialog Act Tagging
[5]	Acoustic Features: Emotional Score by SVM Classifier Linguistic Features: N-gram, Sentiment, and Domain-Specific Words	Automatic	SVM, Max Entropy, Bayesian	Customer Satisfaction Analysis
[6]	Linguistic Features: Dictionary of Feelings	Manual	Logistic Regression	Customer Satisfaction Analysis
[7]	Topic Modelling	Automatic	Decision Tree	Complaint Distribution
[8]	Topic Modelling	Automatic	LDA	Complaint Distribution
[9]	Acoustic Features: MFCC Linguistic Features: N-gram	Manual	SVM DNN	Competitiveness in Conversation

The development of NLP applications has become challenging because computers and humans can communicate through programming languages, which programming languages are usually precise, unambiguous, and highly structured. However, human language is not always precise, often ambiguous and linguistic structures can depend on many complex variables, including slang, regional dialect and social context [11]. Considering the pluralistic Indonesian society and culture of good manners, one of the challenges on Indonesian call center conversation is that there are a lot of non-meaningful words, for courtesy, and the use of non-

standard language or the use of vernaculars and the dialects could decrease the accuracy. It confirms on stated [13] that the challenges of speech are including domain and vocabulary, speaker and accents variability, channels used and noisy audio, natural and spontaneous reaction. Research [14] has investigated Indonesian ASR's performance in recognizing spontaneous speech and read the speech. It resulted that the accuracy decreases 68% on spontaneous speech, as to call center conversations, then read the speech, as news speech and official government speech.

Based on that, it is proposed in the initial study to use text data generated from the manual speech transcribed. It is considered that some specific terms used in telecommunications might not be included yet in the dictionary of existing ASR, and also the raw data from a call center is not sufficiently feasible to directly insert in the ASR tool because of noise and unwanted voices or utterances. In this study, acoustic feature extraction will use MFCC which is commonly used in Speech Recognition [29], but not discussed in detail because the classification system uses only text-based data. But the processing of raw data is planned to produce text and audio data that can be utilized in future research for the construction of a fully automated call center, include develop automatic speech and activity recognition, speaker identifier, even emotion detection.

Another deficiency of previous research is the use of machine learning algorithms instead of applying deep learning. The classification methods used are sufficient to accommodate label targets used in previous studies, but for classification research with a large number of label targets, as our complaint classification which has more than 10 target labels, a more powerful algorithm is needed. The algorithms used are only results-oriented, they result in high accuracy but do not consider the implementation requirement of processing large amounts of data at the same time with fast response. To speed up the process and improve the accuracy deep learning algorithm can be used for classification. Many studies [16-28] have shown that the use of deep learning can increase the accuracy of classification on NLP. From the previous studies, it can be concluded that modifications of CNN and RNN's enhancement most results the higher accuracy than others.

CNN recently achieved very successful results in visual data processing [10]. But with the ability to accommodate feature extraction and classification as one joint task, CNN also could be considered to NLP [28]. RNN is rated as a beneficial algorithm for doing tasks that involve sequential inputs, such as speech and language [10]. RNN is called recurrent because it performs the same task for every element of a sequence, with the output are depended on the previous computations. In other words, RNN has a memory to capture the information about what has been computed so far [15, 30].

Researchers have developed the enhancement model of the original RNN, i.e., LSTM, GRU, and Bidirectional. The memory in Long Short-Term Memory (LSTM) is called cells that take the previous state and the current memory together as input. Internally these cells decide what to keep in and what to erase from memory [11]. It turns out that LSTM cells are very efficient at capturing long-term dependencies. Gated Recurrent Unit (GRU) is similar to LSTM but having fewer gate parameters. Therefore, GRU takes less time for training. Bidirectional comes from the idea that output at a specific time depends not only on the previous element in the sequence but also on future elements [11].

Traditional classifiers based on machine learning methods have defects such as data deficiency, dimension explosion, and poor generalization ability, while classifiers based on deep learning greatly improve these defects, avoid heavy feature extraction process, and have strong learning ability and higher prediction accuracy [31]. Based on the knowledge, in this research, several scenarios for the RNN model algorithm will be proposed for classification.

Besides the differences in the variables of the RNN model, it is also proposed for the action plan to train the model in several different infrastructure environments to determine the speed of training with certain algorithms using the same data.

4 Proposed Works

This section will be explained the steps in developing a classification framework. The developing steps are divided into 3 stages, Stage 1 is building the labeled database, Stage 2 is building the classification model, and Stage 3 is implementation. In this paper stages, 1 and 2 will be presented, while stage 3 is still in the ongoing research phase.

4.1 Building the Database

Fig.2 presents the conscientious approach to produce a database for training with manually transcribing. Data collected from the TELKOM call center record from January to February 2019, the data are expected to represent all problems complained of. Through data obtained from a conversation recording machine, it was found that the metadata format was an 8 kHz / 32bit mono channel audio file. In these recordings, some words are not heard clearly because of the weak audio volume, the noise that was accidentally involved and the piled-up conversation because agents and customers were talking at the same time. To overcome these problems, we look for the best preprocessing action plan so that the audio files can be heard clearly. Through the testing using Google's API for automatic transcripts, the results show that recorded conversations can only be transcribed correctly less than 40%, so sound amplification and noise reduction in the initial audio file is necessary. From several experiments, it was found that to be heard clearly, audio files needed to be amplified at 6 dB and an engineered filter was added for noise reduction. For implementation, it is even strongly recommended to use separated audio channels for agents and customers.
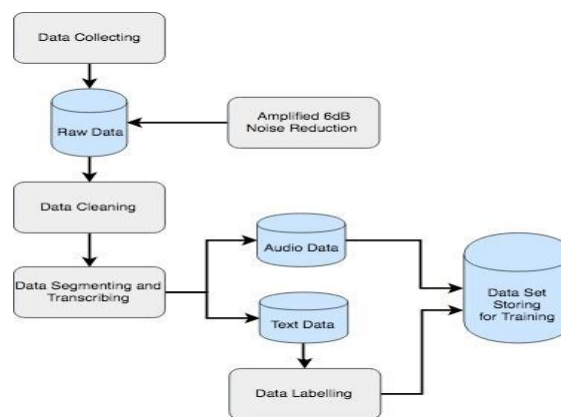

Fig. 2. Dataset preparation workflow.

The data collected are still incomplete conversation records that contain information that needs to be protected, such as customer identity, specific brand (company or product), etc.

Data cleaning has been done by removing sensitive information, the long pause of waiting, meaningless words and unwanted parts of the conversation using audio editor software. Then the speech conversations of the call center will be segmented into sentences and transcribe manually into text. Therefore, the data will be divided into two types: audio data and text data. Both types of data will be processed by feature extraction to result in the automatic speech-to-text generator with good accuracy to be implemented on a real system of automatic complaint classification in the future.

The next step is giving the label of the transcript sentences, according to the activity of the dialogue and the context of customer complaints. Each sentence can have three labels, the problem context, the group of contexts and the responsible parties, but on this research, the sentences will be classified only based on the context of the problem. The results of this stage are labeled database. The data then would be stored in a relational database (MySQL) format and a non-relational database (JSON) format so that it can be used according to system requirements on later implementation.

Table 2. Label context of the sentence.

No.	Group of Contexts	Context
1		Product Information Request
2		Promo Information Request
3		Bills Information Request
4	Information	Registration Information Request
5		Complain Status
6		Provisioning Status
7		Subscription Status
8	Registration	New Subscriber Registration
9		Features Registration
10	Billing Complaint	Bills Payment Failure
11		Bills Do Not Accord
12		All Service Failure
13		Fixed Phone Service Failure
14		Internet Service Failure
15	Technical Complaint	IPTV Service Failure
16		Repetition of Service Failure
17		Hardware Problems
18		Unblocking Service Request
19		Unsubscribe Request
20	Non-Technical Complaint	Blocking Service Request
21		Subscription Package Amendment

No.	Group of Contexts	Context
22		SLG Complain

4.2 Building the Classification Model

In this stage, training data will be carried out to make a classification model using the RNN algorithm. The model was built using the Python programming language and the library of neural network developing tools named Keras. Training will be conducted on 80% of the database produced in the previous stage to produce a model. The remaining 20% of the data will be used in the performance testing of the classification model. The flow of data is shown in Fig.3. Modification of the layers in the RNN will be carried out to result in the expected accuracy.
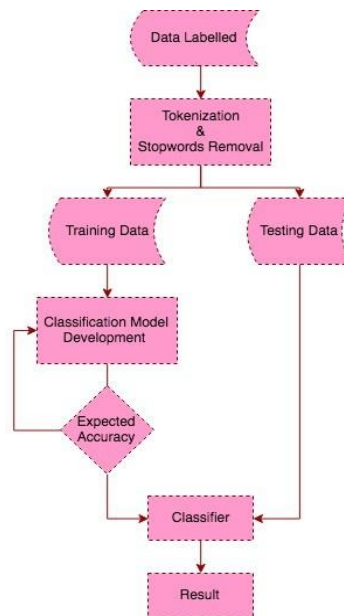
Fig. 3. Classification model development.

In this preliminary research, a simple classification task experiment is performed on the labeled data to review whether the data structure is already correct and can be used in the classification task. Each sentence will be mapped to one of the 22 labels available to indicate what complaint is meant in the statement. The created models are built with 1 LSTM layer and 1 GRU layer on each RNN model. The results are presented in the following table.

Table 3. Classification model training process details.

Model	Epoch	Accuracy (%)	Training Time (s)
LSTM 1	5	72.04	416
LSTM 1	10	85.74	815
GRU 1	5	74.21	318
GRU 1	10	86.61	592

The results of classification still need to be re-evaluated, but it has been able to prove that the deep learning framework with the RNN algorithm can be used to build automatic complaint classification. Further research is still needed to recognize the character of the data and the appropriate model to be used to produce higher accuracy, especially if the expected output is multilabel classification which consists of problem classification labels and labels of responsible parties for example.

5 Conclusion and Future Works

In this paper, the methodological stages that have been carried out are presented, starting from the RC stage, DS-I stage to PS stage. RC stage produces validated research problems and research objectives. DS-I stage, which is done by observing data samples and reviewing the literature, provides the knowledge about research opportunities and suitable algorithms with the parameters that must be considered. PS is done by detailing the research proposal into the model.

However, to develop an effective classification model, it is not enough to rely only on the deep learning network algorithm, it is also important to understand the characteristics of the data and how to prepare it to be appropriate for further processed to achieve the expected classification objectives. So, future research also needs to evaluate the design to produce high accuracy by activating optimal functions in deep learning algorithms. RNN models, including LSTM, GRU and Bidirectional RNN, also various techniques of feature extraction, such as a bag of words and term frequency, which can exploit the information from sentences will be more explored. Common NLP algorithms also will be adopted to enhance the quality of data, because the quality of the input data will affect the accuracy of the system. More data is also needed to enrich the training process, for that reason, raw data can also be retrieved from other customer complaint channels owned by the company, especially text-based channels such as social media, email, or chatbot services.

References

[1] The CallMiner Index Report Q2 2018. Consumer switching by sector, the reasons and the impact of call centers. CallMiner Eureka, 2018.

[2] M. Drama *et al.*, "Malay speech corpus of telecommunication call center preparation for ASR," in *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, Melaka, Malaysia, 2017, pp. 1–6.

[3] Q. B. Nguyen, V. H. Do, B. Q. Dam, and M. H. Le, "Development of a Vietnamese speech recognition system for Viettel call center," in *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*, Seoul, 2017, pp. 1–5.

[4] H. Arsikere, A. Sen, A. P. Pathos, and V. Tyagi, "Novel acoustic features for automatic dialog-act tagging," in *2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 6105–6109.

[5] J. Sun *et al.*, "Information Fusion in Automatic User Satisfaction Analysis in Call Center," in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, 2016, pp. 425–428.

[6] Y. Kobayashi and K. Tsuda, "Extraction of the Customer Satisfaction in the Call Center Using Feelings Dictionary," in *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kumamoto, Japan, 2016, pp. 98–102.

[7] I. O. Yigit, E. Zeydan, and A. F. Ates, "An application for detecting network-related problems from call center text data," in *2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, Istanbul, 2017, pp. 1–4.

[8] F. Kalyoncu, E. Zeydan, I. O. Yigit, and A. Yildirim, "A Customer Complaint Analysis Tool for Mobile Network Operators," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, 2018, pp. 609–612.

[9] S. A. Chowdhury and G. Riccardi, "A Deep Learning approach to modeling competitiveness in spoken conversations," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 5680–5684.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436– 444, May 2015.

[11] M. M. Lopez and J. Kalita, "Deep Learning applied to NLP," *arXiv:1703.03091 [cs]*, Mar. 2017.

[12] L. T. M. Blessing, A. Chakrabarti, and L. T. M. Blessing, *DRM, a design research methodology*. Dordrecht; London: Springer, 2009.

[13] L. Kaushik, A. Sangwan, and J. H. L. Hansen, "Automatic Sentiment Detection in Naturalistic Audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1668–1679, Aug. 2017.

[14] D. P. Lestari and A. Irfani, "Acoustic and language models adaptation for Indonesian spontaneous speech recognition," in *2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, Chonburi, Thailand, 2015, pp. 1–5.

[15] P. Klosowski, "Deep Learning for Natural Language Processing and Language Modelling," in *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, Poznan, 2018, pp. 223–228.

[16] D. Buddhika, R. Liyadipita, S. Nadeeshan, H. Witharana, S. Javasena, and U. Thayasivam, "Domain Specific Intent Classification of Sinhala Speech Data," in *2018 International Conference on Asian Language Processing (IALP)*, Bandung, Indonesia, 2018, pp. 197–202.

[17] S. Ravuri and A. Stoicke, "A comparative study of neural network models for lexical intent classification," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Scottsdale, AZ, USA, 2015, pp. 368–374.

[18] T. Ushio, H. Shi, M. Endo, K. Yamagami, and N. Horii, "Recurrent convolutional neural networks for structured speech act tagging," in *2016 IEEE Spoken Language Technology Workshop (SLT)*, San Diego, CA, 2016, pp. 518–524.

[19] O. Abdelwahab and A. Elmaghraby, "Deep Learning Based vs. Markov Chain Based Text Generation for Cross Domain Adaptation for Sentiment Classification," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Salt Lake City, UT, 2018, pp. 252–255.

[20] S. I. Abudalfa and M. A. Ahmed, "Deep Learning for Target-Dependent Sentiment Classification in Social Media," in *Smart Cities Symposium 2018*, Bahrain, Bahrain, 2018, pp. 40 (5 pp.)-40 (5 pp.).

[21] W. Zhao *et al.*, "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 1, pp. 185–197, Jan. 2018.

[22] J. Cai, J. Li, W. Li, and J. Wang, "Deeplearning Model Used in Text Classification," in *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu. China, 2018, p. 4.

[23] L. Fu, Z. Yin, X. Wang, and Y. Liu, "A Hybrid Algorithm for Text Classification Based on CNN-BLSTM with Attention," in *2018 International Conference on Asian Language Processing (IALP)*, Bandung, Indonesia, 2018, pp. 31–34.

[24] F. Wei, H. Qin, S. Ye, and H. Zhao, "Empirical Study of Deep Learning for Text Classification in Legal Document Review," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 3317–3320.

[25] C. Li, G. Zhan, and Z. Li, "News Text Classification Based on Improved Bi-LSTM-CNN," in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, Hangzhou, 2018, pp. 890–893.

[26] Z. Wang and B. Song, "Research on hot news classification algorithm based on deep learning," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China, 2019, pp. 2376–2380.

[27] N. Abroyan, "Convolutional and recurrent neural networks for real-time data classification," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, Luton, 2017, pp. 42–45.

[28] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative Study of CNN and RNN for Natural Language Processing," *arXiv:1702.01923 [cs]*, Feb. 2017.

[29] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.

[30] T. Tsereteli and D. Knochenwefel, "Application Of Natural Language Processing And Deep Learning Approaches To Natural Language Contained In Sec Filing," presented at the *2018 Eurofiling XBRL Week*, Warsaw, 2018, p. 37.

[31] S. H. Dumpala, I. Sheikh, R. Chakraborty, and S. K. Kopparapu, "Sentiment Classification on Erroneous ASR Transcripts: A Multi View Learning Approach," in *2018 IEEE Spoken Language Technology Workshop (SLT)*, Athens, Greece, 2018, pp. 807–814.

[32] X. I. Quan and J. Sanderson, "Understanding the Artificial Intelligence Business Ecosystem," *IEEE Engineering Management Review*, vol. 46, no. 4, pp. 22–25, Dec. 2018.