How to Cite:
Bisen, W. H., & Agrawal, A. J. (2022). A review on natural language generation. *International Journal of Health Sciences*, *6*(S1), 10365–10376. https://doi.org/10.53730/ijhs.v6nS1.7489

A review on natural language generation

Prof. Wani H. Bisen
Computer Science & Engineering Dept., Shri Ramdeobaba College of Engineering & Management, Nagpur, Maharashtra, India
Corresponding Author email: bisenwh@rknec.edu

Dr. A. J. Agrawal
Computer Science & Engineering Dept., Shri Ramdeobaba College of Engineering & Management, Nagpur, Maharashtra, India
Email: agrawalaj@rknec.edu

***Abstract*---**Natural Language Generation is a field evolved from a computational linguistic, the discipline concerned with understanding written and spoken words and building artifacts that usually process and produce language. The emphasis of NLG is on computer systems that can produce understandable texts in human languages. It is one of the fastest growing applications of Artificial Intelligence as it articulately communicates ideas from data at remarkable scale and accuracy. NLG includes variety of application areas such as Healthcare, Finance, Human Resources, Legal, Marketing, Sales, Operations, Strategy, and Supply Chain. The field of NLG has changed drastically in last few years with the emergence of successful deep learning methods. This paper focuses on deep learning techniques and methods used for Natural Language Generation by reviewing some of the recent work done in this direction, mainly in the field of healthcare. In recent times the need for automatic medical report generation has become a new challenge for the doctors and physicians, so that they could spend quality time with patients rather than investing time in preparing manual reports. In this paper we will discuss the performance of statistical language generator based on semantically controlled LSTM structure, Context-aware NLG with RNN, Gated Attention Neural Networks (GANN), Named Entity Recognition (NER) model, RNN-LSTM generative model for report generation. This can be helpful to develop a deep learning model that will be best suitable for automatic medical report generation.

***Keywords*---**Natural Language Generation (NLG), Long Short-term Memory (LSTM), Recurrent Neural Network (RNN), Gated Attention Neural Networks (GANN).

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 9 May 2022

Introduction

Natural language generation provides computer ability to generate a language which is easily understandable by human. In healthcare, it plays an important role in providing effective communication between physicians and their patients. Medical Transcription Process has been introduced in recent years which involve transcribing the recorded audio files dictated by physicians or other healthcare professionals into text. Transcription is the process of converting the audio recording into formatted text. It is an integral part of creating clinical documentation. In transcription process, Physicians dictate notes that are written out by a medical transcriptionist and reviewed by quality assurance personnel. The physician reviews, modifies, and signs the preliminary report, which is then delivered to those who need to see it. Physicians prescribe hundreds of millions of reports each year. The transcription industry has only recently begun using the Internet, and many hospitals and clinics are lagging behind modernizing their infrastructure to support automation of more processes.

In recent years, home based medical transcriptionist has become a lucrative option in transcription industry but most medical transcriptionist are not healthcare professionals, therefore there is a chance of erroneous medical report, also there is always a risk of health information of the patient being stolen, lost or misused therefore, there is a need of an automated transcription system which can convert an unstructured text into the structured repot by using deep learning techniques for natural language generation. In this paper we will analyze different deep learning techniques used for natural language generation. We will propose a model based on the reviews and performance analysis of the different deep learning techniques and NLG methods that are used for language generation and preparation of structured report. Applications of NLG methods includes generation of various reports such as weather reports, patient reports, profit and loss reports, chatbots, generating expert system explanation and critiques, content creation

Literature Review

Authors Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke and Steve Young suggested a general statistical language generator based on a semantically controlled short-term storage structure (LSTM) in their document entitled "Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems”. The conventional approach to NLG was based on sentence planning and surface realization. Sentence planning is the method to choose the words and structures to fit information into sentence sized_unit while surface realization is about determining the surface form of output. The author in their paper aimed to use a neural network based approach for language generation. RNN can generate significant sentences by learning from the character_level corpus, but it is difficult to form a RNN model with long-term dependencies due to gradient vanishing problem. Long Short-term Memory (LSTM) network overcomes vanishing gradient problem. The gradient vanishing problem includes sigmoid activation in the RNN recurring connection with an auto current memory block and a set of multiplication gates for copying playback, writing and resetting the operations in digital computers. Long Short-term

Memory (LSTM) Network overcomes this problem of gradient vanishing. The authors have also proposed a Neural Language Generation Model which is based on RNN architecture and Semantically Controlled LSTM Cell whose upper part work for surface realization and lower part work for sentence planning based on sigmoid control gate and dialogue act (DA).

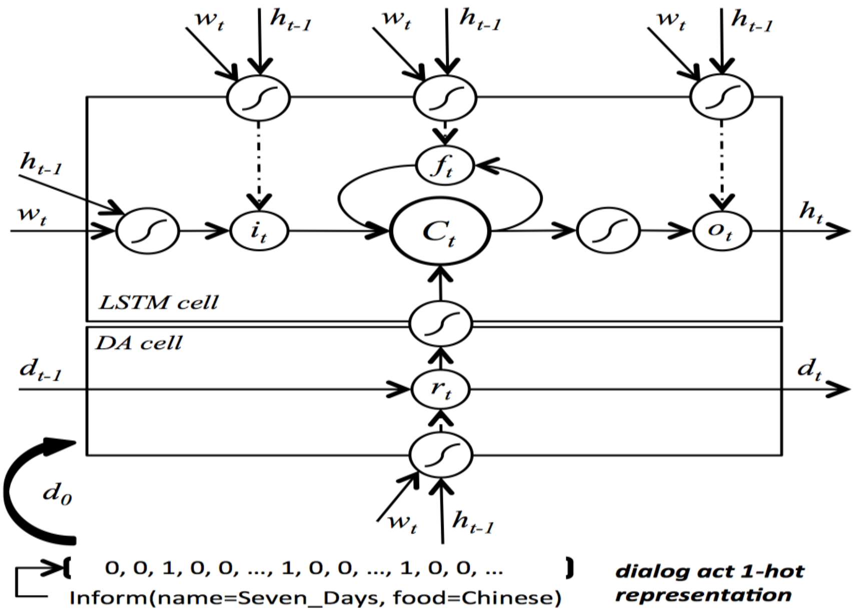

Fig. 1 Semantically Controlled LSTM Cell

The proposed architecture is defined by following equations:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{wi}\mathbf{w}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1}) \qquad (1)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{wf}\mathbf{w}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1}) \qquad (2)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{wo}\mathbf{w}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1}) \qquad (3)$$

$$\hat{\mathbf{c}}_t = tanh(\mathbf{W}_{wc}\mathbf{w}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1}) \qquad (4)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t \qquad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot tanh(\mathbf{c}_t) \qquad (6)$$

where б is the sigmoid function, $i_{t.}$, o_t, $f_t \in [0,1]^n$
are input, output and forget gates respectively, and c_t' and c_t are proposed cell value and true cell value at time t. The authors have extended the neural language generator model to be deep in both time and space by stacking multiple LSTM Cells on top of the original structure.

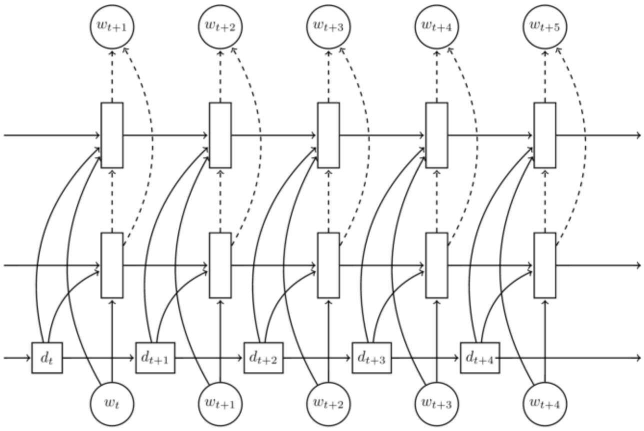

Fig. 2 Deep LSTM Generator

The corpus is collected and used for training, testing and validation phase in 3:1:1 ration. To show the scalability of the proposed method and model performance the authors have experimented on two areas that talk about restaurants and hotels respectively. The objective evaluation of the five major outcomes is as follows.

Table 1
Objective Evaluation

Method	SF Restaurant		SF Hotel	
	BLEU	ERR(%)	BLEU	ERR(%)
hdc	0.451	0.0	0.560	0.0
kNN	0.602	0.87	0.676	1.87
classlm	0.627	8.70	0.734	5.35
rnn w/o	0.706	4.15	0.813	3.14
lstm w/o	0.714	1.79	0.817	1.93
rnn w/	0.710	1.52	0.815	1.74
lstm w/	0.717	0.63	0.818	1.53
sc-lstm	0.711	0.62	0.802	0.78
+deep	**0.731**	**0.46**	**0.832**	**0.41**

The Deep SC-LSTM is significantly better than the rnn w , rnn w/0 and knn. The neural network based language processing can be useful as an implicit distributed presentation of words and single compact parameter encoding of information to be conveyed. The crucial implicit advantage of neural network grounded language processing is the implicit use of distributed representations for words and a single compact parameter encoding of the information to be conveyed. It's possible to further condition the creator on some dialogue features. This type of the converse information during the discussion by espousing corpus grounded governance enables sphere scalability and multilingual NLG to be achieved with lower cost and a shorter lifecycle.

Authors Jian Tang, Yifan Yang, Sam Carton, Ming Zhang, and Qiaozhu Mei have suggested two novel approaches to encode the context into a continuous sematic representation and to decode the semantic representation into text sequence by using Recurrent Neural Network, in their document entitled “Context-aware Natural Language Generation with Recurrent Neural Networks” have proposed. The Proposed model aimed to generate not only semantically and syntactically coherent sentences, but also the sentences that are reasonable at particular contexts. Indeed, contexts have been proved to be very useful for various natural language processing tasks such as topic extraction, text classification and language modeling. They have proposed two novel approaches for context-aware natural language generation, which map a set of contexts to text sequences. In the first model C2S encodes a set of surrounds into a nonstop representation and also crack the representation into a textbook sequence through a intermittent neural network.

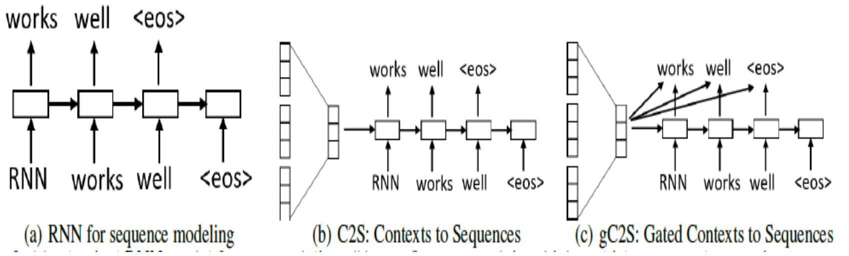

Fig. 3 (a) classical RNN model for text modeling; (b) first approach in which multi-layer neural networks are encoder and RNN are decoders; (c) second approach which adds skip-connections from contexts to words, controlled by a gating function

They have evaluated the approaches on the user reviews data. Experimental results show that further than 50% of the fake reviews generated by our approach are misclassified by mortal judges, and further than 90% of the reviews are misclassified by being fake review detection algorithm. In the future, they plan to integrate more context information, e.g., the user, the detailed descriptions of the products, the product prices, into their approaches and also evaluate the approaches in other scenarios like generating the titles of scientific papers based on the author, venue, and time information. It may be also beneficial to improve model through the attention mechanical i.e., attending to different types of contexts when generating words in different positions.

Authors, Hai-Tao Zheng, Wei Wang, Wang Chen, and Arun Kumar Sangaiah have proposed a gated attention neural network model to generate news comments in their paper titled “Automatic Generation of News Comments Based on Gated Attention Neural Networks”. The GNNN model use the news content self-adaptively and selectively. For the Proposed model the authors have collected the news titles $x = x_1, x_2, x_3 \ldots x_m$ an inputs and the model needs to generate relevant news comments $y = y_1, y_2, y_3 \ldots . y_m$ as an output. The proposed model contains two components: comment generator and comment discriminator. The comments generator converts all words $x = x_1, x_2, x_3 \ldots x_m$ titles into one-hot vectors and obtains the embedding representations by multiplying the embedding matrix. The embedding representation of words is computed as follows.

$$\vec{e}_i = E\vec{x}_i,$$

The Comment decoder converts sequences of comments words y = $y_1, y_2, y_3.....y_m$ into one hot vector and gets their low-dimensional representations through the shared embedding matrix E.

$$\vec{e}_j = E\vec{y}_j.$$

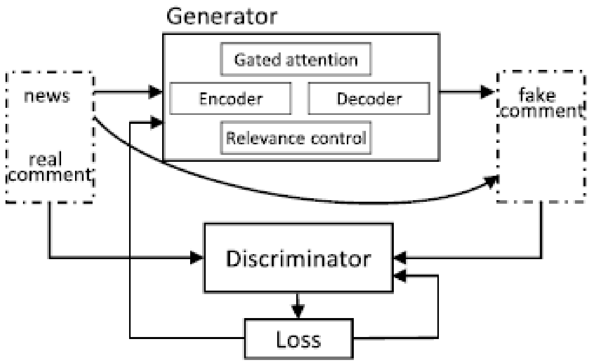

Fig. 4 Gated Attention Neural Network Model

We need to handle the contextual information efficiently to guarantee the contextual relevance between comments and news. In this paper the authors have also tackled the problem of dependency of words on proceeding words and not on the context. This problem may be tackled through gated attention mechanism which allows a selective gate to select contextual information whenever needed.

Table 2
Evaluation results on the test set

Method	Perplexity
Enc2Dec	196.6
Enc2Dec+ga	168.7
Enc2Dec+ga+rc	162.6
GANN	162.1

From the above table it is observed that result of basic encoder and decoder is worst while other methods taking account of contextual information are better. The gated attention neural network deals with the new context effectively and improves the performance significantly. Other methods like planning model can be integrated to this model. To improve the performance further we can combine the word representation with character representation.

The Authors, Cheng-Tse Wu, Hsiao-Ko Chang, Ji-Han Liu, Jyh-Shing Roger Jang have proposed NER model to generate a structured medical report using machine learning technique in their paper titled “Adaptive Generation of Structured Medical Report using NER regarding Deep Learning. The authors have aimed to propose a Named Entity Recognition (NER) Model for chest X-ray medical report generation.

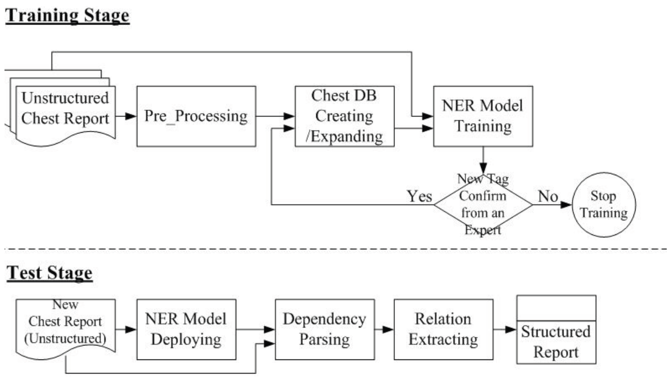

Fig. 5 NER Model for Chest X-ray report

This model consists of two stages: 1. Training Stage 2. Testing Stage. During the training phase, the unstructured electronic medical record is obtained from physicians. After preprocessing this file, the database is generated for the contents of the chest report data and we conduct NER training and modeling using NeuroNER. We determine if the new chest X-ray labels are available or not to repeat the NER training and modeling step to generate a new model. In the test phase, we deploy to test the new chest report and perform dependency analysis and relations extraction to generate the structured report.

Table 3
Test Result of NER Model

Our Method					
n=9,933	predicted		Precision	Recall	F1-Score
observed	TP=5,507	TN=67	98.80%	80.65%	88.80%
	FN=1,321	FP=3,038			

From the test results the authors have observed that the recall value is lower than perfection value. In order to ameliorate the value of recall the authors have anatomized all structured affair reports classified into FN and determined three types of crimes similar as Keywords not being in the database, Typo error, Lower-case acronym or abbreviation.

Authors, Assaf Hoogi, Arjun Mishra, Francisco Gimenez, Jeffery Dong, and Daniel L. Rubin suggested RNN-LSTM architecture to generated mammography reports by preserving style and contents of the report, in their document entitled “Natural Language Generation Model for mammography reports simulation”. The authors have presented LSTM- RNN architecture to simulate the realistic mammography reports.

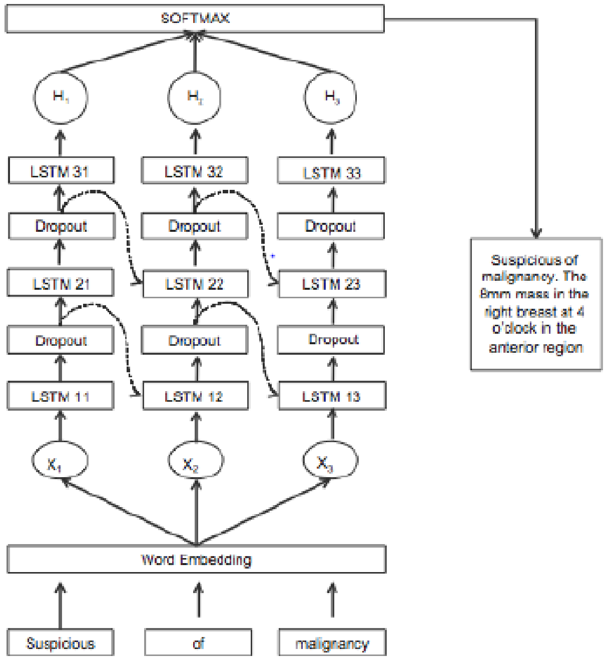

Fig. 6 Architecture for text-to-text report generation (RNN-LSTM)

This model generates the result sequentially where the next scheduled item is conditioned or unconditioned on the previously generated items. Conditional generative text-to-text RNN-LSTM model maps the input sequence ($x_1 \ldots x_T$) into an output sequence $y_{1 \ldots} y_{T'}$. The authors have used word embedding method to capture both semantic and syntactic variations of words. They have used GloVe (Global Vectors for Word Representation) mode which is an unsupervised learning algorithm for obtaining vector representations. To train GloVe model an iterative learning is done through sampling the word co-occurrence distributions. The LSTM architecture was train by using the whole MCW cohort.

Table 4
Multilevel Statistics of Mammography Reports

Parameter	Data type	Value
Word length	MCW - Real Stanford - Real RNN - Simulated	5.72 ± 0.11 6.06 ± 0.24 5.5 ± 0.31
Words in a sentence	MCW - Real Stanford - Real RNN - Simulated	12.2 ± 1.09 16.2 ± 1.48 14.8 ± 2.34
Words in a report	MCW - Real Stanford - Real RNN - Simulated	235.6 ± 48.9 231.2 ± 134.6 225.8 ± 66.5

For testing 27 most frequent keywords were selected for each report. Among those, 24 keywords were identical between the real and the simulated reports

(0.889 of the frequent words). The performance and the accuracy of the machine learning model can be improved further that requires large amount of clinical text as their major input.

Proposed Methodology

In this paper we are proposing the following model to generate structured report from the audio file given as an input to the speech to text converter. After converting the spoken words into text we will preprocess the text to remove all the noise and inconsistencies from the converted text. We will design a deep learning model by integrating best Neural Network architectures such as RNN, LSTM etc. and will train, test and validate our model for better performance and accuracy check.

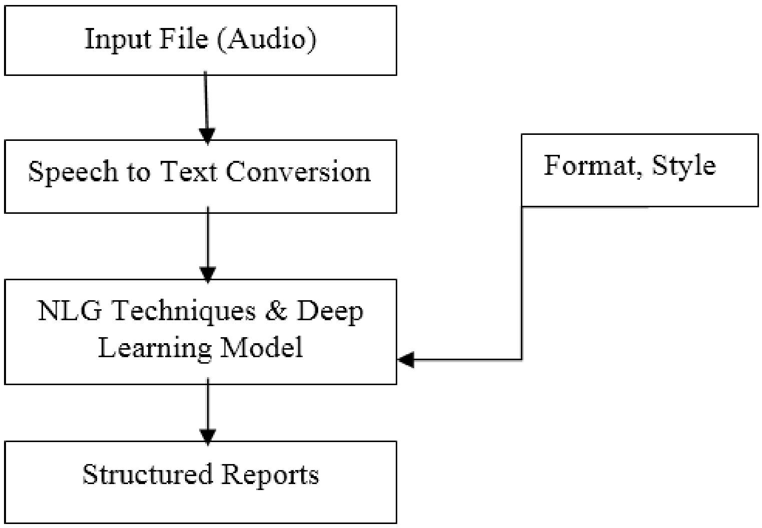

Fig. 7 Flow chart of Proposed Model

1. Speech to Text Converter: The input audio file is converted into text by using any speech to text converters such as Amazon Transcribe, IBM Watson.
2. Deep Learning Model: The deep learning models such as RNN, LSTM, Bi-directional LSTM can be used to train the dataset to generate the concise report.
3. Structured Report Generation: A Well formatted structured medical report can be obtained from the proposed method which can be helpful for the physicians to streamline patient’s treatment.

Research GAP

This review paper is helpful in understanding the research gap present in the area of natural language generation which is the sub filed of natural language processing. These Research gaps are as follows:

- Need of adopting a corpus based regime which enables domain scalability and multilingual NLG to be achieved with less cost and a shorter lifecycle.
- Need of attending to different types of contexts when generating words in different positions to improve the model.
- Need to integrate other methods like a planning model to generate longer statements such as comment articles.
- Need to incorporate user information to generate personalized comments. Moreover, we will combine word representation with character representation to improve the performance.
- Need of optimization and accuracy improvement.
- Need of improving the classification accuracy of a variety of machine learning models that require large amounts of clinical text as their major input.

Conclusion

In this review paper we have tried to understand the work done in the field of Natural Language Generation. We have tried to uncover all the research gaps which can be filled further by integrating the deep learning models such as RNN, LSTM, Bi-directional LSTM. We will choose the best model which is suitable for classification, analysis and prediction. We will develop a model to generate structured reports from unstructured input data (spoken words). We will try to achieve highest accuracy and performance by saving time and cost spent in medical transcription process.

Future Scope

In future we will try to fill the entire Research gap. The structured report can be used to generate automatic health records to facilitate patient outcomes and streamline the treatment of the patients by healthcare providers. These medical reports can also be useful to share medical records for the further studies which help to promote methodological research shared for the further studies which helps to promote methodological research.

References

[1] Tsung-HsienWen, Milica Ga˘si´c, Nikola Mrk˘si´c, Pei-Hao Su, David Vandyke and Steve Young, "Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems". August, 2015, in publication of Researchgate.net. https://doi.org/10.48550/arXiv.1508.017

[2] Jian Tang, Yifan Yang, Sam Carton, Ming Zhang2, and Qiaozhu Mei , "Context-aware Natural Language Generation with Recurrent Neural Networks," 29th Nov 2016 in publication of Researchgate.net. https://doi.org/10.48550/arXiv.1611.09900

[3] HAI-TAO ZHENG, WEI WANG , WANG CHEN, AND ARUN KUMAR SANGAIAH," Automatic Generation of News Comments Based on Gated Attention Neural Networks," 14 Feb, 2018, Publication in IEEE Access , Special section on Recent Computational Methods in Knowledge Engineering and Intelligence Computation.

DOI.10.1109/ACCESS.2017.2774839
https://ieeexplore.ieee.org/abstract/document/8114175

[4] Cheng-Tse Wu, Hsiao-Ko Chang, Ji-Han Liu, Jyh-Shing Roger Jang," Adaptive Generation of Structured Medical Report using NER regarding Deep Learning," *2018 IEEE Conference on Technologies and Applications of Artificial Intelligence*. DOI 10.1109/TAAI.2018.00012. https://ieeexplore.ieee.org/abstract/document/8588468

[5] Assaf Hoogi, Arjun Mishra, Francisco Gimenez, Jeffery Dong, and Daniel L. Rubin, "Natural Language Generation Model for mammography reports simulation ," 29 May, 2020 Computation and Language (cs.CL); Artificial Intelligence (cs.AI), DOI 10.1109/JBHI.2020.2980118.

[6] Kumar, Vijay and Singh, Hemant and Mohanty, Animesh, "Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation and Abstract Meaning Representation" (April 3, 2020). International Journal of Engineering and Advanced Technology (IJEAT) , Volume-9 Issue-4, April, 2020, Page no:2361-2365, Available at SSRN: https://ssrn.com/abstract=3925035.

[7] V. Ramesh and A. Kolonin, "Natural Language Generation Using Link Grammar for General Conversational Intelligence," 19th April 2021 Computation and Language (cs.CL); Artificial Intelligence (cs.AI),arXiv:2105.00830 [cs.CL].

[8] N. Sadman, M. H. Rahman, S. Tasneem, M. A. Haque and K. D. Gupta,"Recommend Speciality Doctor from Health Transcription: Ensemble Machine Learning Approach," *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, 2021, pp. 0967-0972, doi: 10.1109/CCWC51732.2021.9376111.

[9] M. Xu, M. Islam, C. Ming Lim and H. Ren, "Learning Domain Adaptation with Model Calibration for Surgical Report Generation in Robotic Surgery," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 12350-12356, doi: 10.1109/ICRA48506.2021.9561569

[10] M. Zhang, Y. Zhou, L. Zhao and H. Li, "Transfer Learning From Speech Synthesis to Voice Conversion With Non-Parallel Training Data," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1290-1302, 2021, doi: 10.1109/TASLP.2021.3066047

[11] Y. Yamazaki, Y. Chiba, T. Nose and A. Ito, "Filler Prediction Based on Bidirectional LSTM for Generation of Natural Response of Spoken Dialog," *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 2020, pp. 360-361, doi: 10.1109/GCCE50665.2020.9291867

[12] M. D. Marin, E. Mocanu and C. Seifert, "Effectiveness of neural language models for word prediction of textual mammography reports," *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 1596-1603, doi: 10.1109/SMC42975.2020.9283304

[13] Sashank Santhanam, Samira Shaikh "A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past, Present and Future Directions", 2019, journal ArXiv, volume abs/1906.00500.

[14] R. Wang, Z. Li, J. Cao, T. Chen and L. Wang, "Convolutional Recurrent Neural Networks for Text Classification," *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1-6, doi: 10.1109/IJCNN.2019.8852406

[15] H. Yeo, "A Machine Learning Based Natural Language Question and Answering System for Healthcare Data Search using Complex Queries," *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 2467-2474, doi: 10.1109/BigData.2018.8622448.

[16] Lee, S.H. "Natural language generation for electronic health records". Centers for Disease Control and Prevention, Atlanta, GA, USA, *npj Digital Med* 1, 63 (2018). https://doi.org/10.1038/s41746-018-0070-0.

[17] C. -T. Wu, H. -K. Chang, J. -H. Liu and J. -S. R. Jang, "Adaptive Generation of Structured Medical Report Using NER Regarding Deep Learning," *2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, 2018, pp. 10-13, doi: 10.1109/TAAI.2018.00012

[18] Burhanuddin Lakdawala, Farhan Khan, Arif Khan "*Voice to Text transcription using CMU Sphinx*"Proceedings of the 2nd International Conference on InventiveCommunication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2.

[19] Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao, "*Recurrent Convolutional Neural* Networks for Text Classification", *2015* Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

[20] R. Vaish, S. T. Ishikawa, J. Liu, S. C. Berkey, P. Strong and J. Davis, "Digitization of health records in rural villages," *2013 IEEE Global Humanitarian Technology Conference (GHTC)*, 2013, pp. 209-214, doi: 10.1109/GHTC.2013.6713682.

[21] Y. Zhao, "Speech-recognition technology in health care and special-needs assistance [Life Sciences]," in *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 87-90, May 2009, doi: 10.1109/MSP.2009.931993

[22] B. Zhou, Y. Gao, J. Sorensen, Z. Diao, M. Picheny, "Statistical Natural Language Generation for Speech-to-Speech Machine Translation Systems", ICSLP-2002:Inter. Conf. on Spoken Language Processing, vol. 3, pp. 1897-1900, Denver, CO USA, Sept. 2002.

[23] Jagannathan, V. "*The Careflow architecture. A case study in medical transcription". ,* (2001). *5(3), 0–64.* doi:10.1109/4236.935178

[24] F. Vorbeck, A. Ba-Ssalamah "*Report generation using digital speech recognition in radiology"* 2000, Springer-Verlag DOI:10.1007/s003300000459 Corpus ID: 2838316

[25] Reiter, E., & Dale, R."*Building Natural Language Generation Systems"* (Studies in Natural Language Processing). (2000). Cambridge: Cambridge University. Press.doi:10.1017/CBO9780511519857.