

Automated Disease Prediction and Prescription System with Regional Kannada Synonym Integration

Giridhara Gouda

giridhargowda76@gmail.com

Sri Siddhartha Academy of Higher Education

Suma R

Sri Siddhartha Academy of Higher Education

Research Article

Keywords: Robot, Machine Translation, Recurrence Neural Networks (RNN), Google translator, Whisper translator

Posted Date: September 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10827458/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Automated Disease Prediction and Prescription System with Regional Kannada Synonym Integration

Giridhara Gouda¹[0000-0002-1379-7229] and Suma R²[0009-0001-9510-6291]

¹ Sri Siddhartha Academy of Higher Education
Tumkur ,Karnataka, India.
giridhargowda76@gmail.com

² Sri Siddhartha Academy of Higher Education
Tumkur ,Karnataka, India
sumar@ssit.edu.in

Abstract: Patients can converse their diseases in natural language to hospital robots, which then examine the input to forecast and applaud the resultant medication. This paper covers a machine translation model that exploits Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units to translate Kannada text into English. A significant feature of the proposed model is its capability to predict localized medical synonyms for numerous diseases—a method where current approaches like Google Translator frequently fail due to regional language disparities. By appropriately recognizing disease names through these regional synonyms, the system endorses suitable treatments, serving healthcare professionals manage high patient volumes more proficiently. When assessed against Google Translate and Whisper, the anticipated RNN-LSTM model validated superior performance in translation accuracy and context recognition.

.Keywords: Robot, Machine Translation, Recurrence Neural Networks (RNN), Google translator, Whisper translator.

1 Introduction

Superior medical care depend deeply on vibrant communication between healthcare providers and patients. Nevertheless, language barriers frequently disrupt this dialogue, mainly in diverse and multilingual communities. When providers and patients cannot understand each other, it ominously escalates the risk of misdiagnoses, medication errors, and complete inferior patient outcomes [1]. The critical prominence of healthcare has powered a sharp increase in the need for medical speech translation [2], which plays a vibrant role in supports migrants and international patients traveling for particular treatments [3]. By breaking down language hurdles, speech translation (ST) confirms that the quality of patient care is never conceded [4]. At the same time, the improvement of highly proficient, domain-adaptable pre-trained prototypes has glowed central research attention in this field. However, moving these revolutions into real-world use is deeply delimited by a absence of public datasets and models, largely due to stern data

privacy protocols[5] Most present public datasets for medical machine translation (MT) are limited to text, lack scale, and are frequently scraped from fragmented web sources. Additionally, past research in medical ST has classically absorbed on narrow software applications rather than a methodical, research-driven approach [6].

Speech translation is fundamentally the process to convert the language spoken by a person into another desirable language. Google translator uses an algorithm that implements Artificial Neural Network (ANN) to expedite machine translation with Neural Machine Translation (NMT), termed GNMT (Google NMT). Now, Recursive Neural Network (RNN) takes sequence of data, like time series, so that individual sample reliant on the former ones, unlike Artificial Neural Network. Artificial neurons processes by accepting input signals, processing them, and transferring output to connected nodes. Unlike typical Artificial Neural Networks (ANNs), **Recurrent Neural Networks (RNNs)** outshine at exhibiting sequential data such as time series by managing dependencies between successive samples. Additionally, RNNs can be combined with convolutional layers to multiply the active pixel neighbourhood in spatial tasks. Due to continuously uses back propagation over time in memory cells, RNNs can link general temporal gaps while strongly manages noise and continuous data [7]. Lastly, unlike Hidden Markov Models (HMMs), RNNs do not need a predefined volume of states, permitting them to dynamically accomplish an infinite state space

Current research designates that Neural Machine Translation (NMT) systems overtake traditional translation approaches in learning speed while providing smoother, more accurate results. Within NMT, Recurrent Neural Network (RNN) architectures—precisely those employing Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) designs—are extremely popular because they outshine at following the natural flow and consecutive order of words. Awkwardly, despite these technological progressions, very slight consideration has been given to generating translation tools for low-resource languages, predominantly those spoken across Tulu, Konkani and other Indian regional languages [8]. To bridge this research breach, this paper modelled an NMT system devoted to translating Kannada into English with different synonyms.

Our key contributions are summarized as follows:

- We introduce primarily optimistic translation system for the Kannada–English language pair, integrating attention mechanisms [9] and universal consideration [10].
- We incorporate the teacher forcing algorithm to deliver the decoder and encoder with precise reference outputs in training, so refining translation quality.
- We developed and compared both Greedy Decoding and Beam Search approaches to boost translation fluency and accuracy.
- We have conducted broad experiments with approximately 900 equivalent sentence pairs or datasets. This transformer-based transfer learning system [11, 12, 13] attains a ultimate translation accuracy of 90% and validates a prominent development of +4.83 BLEU points over existing benchmarks for Kannada–English translation

- Developed better translation model to translate different words with same meaning (synonyms) to identify proper disease.

The rest of this paper is organized as follows. We originate by studying pertinent literature in NLP and NMT in Section 2. Section 3 presents the foundational ideas of RNNs and summaries their intrinsic limitations. In Section 4, we point the architecture of our planned translation model. Section 5 designates the experimental outline, shows RNN variants, attention mechanisms, decoding policies, datasets, and transfer learning applications. Our quantitative and qualitative results, together with performance assessments, are shown in Section 6. Finally, Sections 6 and 7 provide a complete discussion, conclude the paper, and summarizes limitations as well as possibilities for future research.

2. Related Work

The roots of Machine Translation (MT) came from 1946, when Warren Weaver projected the fundamental idea of giving foreign languages like secret codes that could be splintered using cryptography. Over the decades, the field changed by mixing two different methods: rule-based logic and data-driven statistical approaches. This synthesis ultimately covered the way for today's radical hybrid models in Neural Machine Translation (NMT) [14].

Modern Neural Machine Translation (NMT) usually relies on a sequence-to-sequence (Seq2Seq) basis, which contains two key components: an encoder and a decoder. The encoder compresses the source sentence into a single, fixed-size vector holds its meaning, and the decoder customs this vector to build the translated text. Nevertheless, typical encoder-decoder frameworks suffer from two major limitations. First, long-range associations between words frequently deteriorate, which reduces translation quality; Gusev and Oboturov opposed this by retreating the encoder's processing path through training to recover context retention. Second, squeezing a complete sentence into a private vector creates an information block, a defect that becomes particularly severe with longer inputs.

By familiarizing an responsiveness mechanism, Bahdanau[15] bypassed the confines of a single, fixed-size vector. This method permits the decoder to dynamically takes all encoder states, indicate the precise input tokens desirable for each step of the translation. While this radically boosted translation accuracy, it presented advanced computational overhead. Additionally, the model still looks challenges when handling rare training data or taking very long-term dependencies.

To improve token arrangement, Luong[16] consequently implemented a native attention mechanism. Constructing these progressions, self-attention networks—familiarised by scientists like Yang[17] and Shaw [18]—ominously strapped sequence modelling forward by permitting tokens to interact with and assess various positions through the whole sentence.

Apart from recurrent architectures like LSTMs and GRUs, Convolutional Neural Networks (CNNs) have developed as a feasible for neural machine translation. Early work by Kalchbrenner[19] and Bradbury[20] used CNNs to attain extremely efficient

sequence modelling, while Gehring[21] later incorporated attention mechanisms with CNNs to efficiently detect long-range dependencies. Furthermore, sequence-to-sequence translation has been boosted by other architectural revolutions, such as the Connectionist Temporal Classification (CTC) structure presented by Graves[22] and the use of ensemble learning methods to maximize prediction process [23].

Neural machine translation qualified a key model with the introduction of the Transformer mechanism by Vaswani[24]. By totally removing recurrent structures in favor of self-attention frameworks, this technique allowed for parallelization through training while concurrently increasing translation quality. The architecture depending on vital building blocks such as multi-head attention and position-wise feed-forward layers. Eventually, the Transformer aided as the prime basis for subsequent state-of-the-art models, with GPT-2 [25], BERT [26], and T5 [27].

While conventional NLP backgrounds depend deeply on sequential or convolutional designs, modern research suggests that Graph Neural Networks (GNNs) [28] shines at capturing the structured, non-linear relations that present in language, including syntactic and semantic dependencies [29]. Scientists have successfully organized GNNs for various applications, such as word embedding generation [30], knowledge graph completion, and machine translation in low-resource backgrounds, principally because they can map long-distance interactions among textual elements within a graph system [30, 31]. In translation workflows exactly, mapping words as nodes while treating their syntactic co-occurrences as edges permits the system to retain a ample deeper layer of contextual information.

Additionally, the arrival of Large Language Models (LLMs)—with the GPT family, multilingual BERT schemes, and modern generative embedding approaches—has ominously advanced the limits of machine translation and cross-lingual knowledge. By using massive-scale pre-training, these architectures comes extremely contextualized representations, ensuring strong enactment even when trade with low resonance languages [32, 33]. Presently, generative embedding are available in cross-lingual information retrieval, text summarization, and neural translation [34, 35].

This research work focuses on translating between Kannada to English with synonyms usually comes in Kannada language. To accomplish this, we deploy a typical RNN-based encoder-decoder architecture, amplified with both general and global attention approaches [36, 37]. Moreover, we developed the teacher forcing algorithm [38] to uses the decoding point through model training, while adopting transfer learning techniques to optimize entire translation efficiency.

3 Recurrence Neural Networks (RNN) Architecture

Designed precisely for temporal and sequential data, Recurrent Neural Networks (RNNs) uses internal hidden states to recall historical information through successive time steps. This sequential dispensation ability makes them extremely effective for time-sensitive real time applications such as financial forecasting, electrical load prediction, and water quality monitoring. Nevertheless, original architectures like the Vanilla RNN suffering from intrinsic vulnerabilities during back propagation through time

(BPTT), the vanishing and exploding gradient phenomena. These constraints deeply limit the model's accuracy to retain long-term dependencies, initiating standard RNN accuracy to reduce severely as sequence lengths expand [39, 40]. Figure 1 depicts the architecture of a standard Recurrent Neural Network (RNN), which computes sequential data by using the similar weight matrices (U , V , and W) at each time step

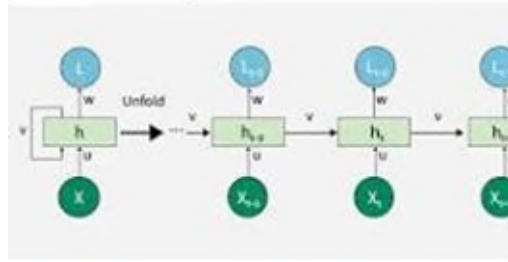


Fig. 1. Standard (Vanilla) RNN architecture. The weight matrices U , V , and W are shared through time steps to compute temporal sequences

To address these confines, more refined architecture of Long Short-Term Memory (LSTM) A **Long Short-Term Memory (LSTM)** network is a progressive variant of Recurrent Neural Networks (RNNs) precisely designed to resolve the vanishing and exploding gradient constraints. While typical RNNs overwrite their hidden states entirely at each step, LSTMs familiarize an internal cell state (C_t) treats as a conveyor belt, together with a sophisticated gating mechanism that normalizes the flow of information. The figure 2 depicts the same.

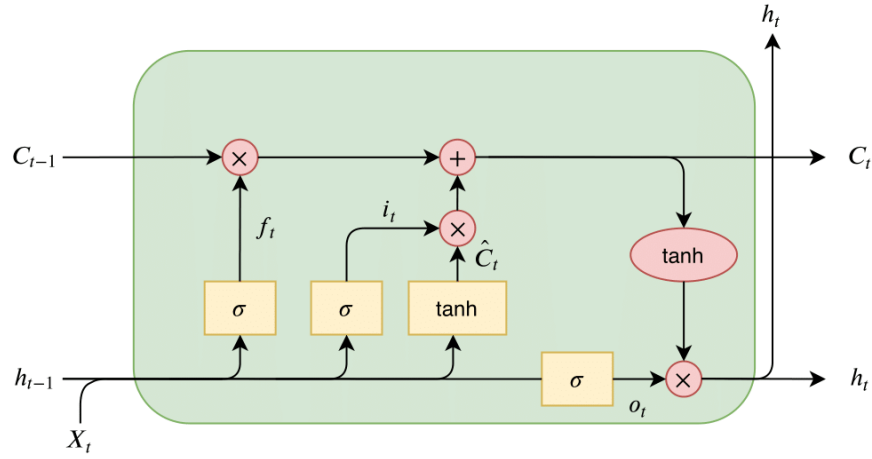


Figure 2. Overview of LSTM Architecture

At any given time step t , an LSTM unit takes the present input X_t and the preceding hidden state h_{t-1} , processing them over three principal gates:

1. The Forget Gate (f_t)

Identifies what information from the preceding cell state C_{t-1} should be discarded or retained

$$f_t = \alpha(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2. The Candidate Memory (i_t) and Input Gate ($\bar{C}_t$)

Selects what new data from the present input should be stored in the cell state

$$i_t = \alpha(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

3. Cell State Update

Merges the old cell state with the new candidate memory to get the rationalized long-term memory.

$$C_t = f_t \odot C_{t-1} + i_t \odot \bar{C}_t \quad (4)$$

4. Hidden State (h_t) and Output Gate (O_t)

Identifies parts of the updated cell state becomes output as the new hidden state.

$$O_t = \alpha(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t \odot \tanh(C_t) \quad (6)$$

Beyond speech recognition and machine translation, LSTMs are extensively recognized for their strength in numerous additional tasks. For example, researchers combined Word2Vec with an LSTM architecture to perform sentiment analysis, achieving a remarkable accuracy of 89.96% and emphasizing the model's excellent efficiency in natural language processing. Additionally, the formation of various LSTM variants endures to broaden its real-world use cases through new domains. To diminish the vanishing gradient constraint common in typical RNNs, Hochreiter and Schmidhuber [42] presented Long Short-Term Memory (LSTM) networks. The central invention of the LSTM is its use of gating mechanisms—specially input, forget, and output gates—which regulates how data moves through an internal memory cell. By selectively keeping or neglecting data across long sequences, LSTMs outshine at memorizing long-term dependencies. The applications such as speech processing and handwriting recognition, LSTM networks have recognized state-of-the-art routine. Additionally, when paired with the Connectionist Temporal Classification (CTC) loss function, they can be efficiently trained for applications that request flexible sequence alignment. These

LSTM-driven frameworks have attained benchmark-setting outcomes in global pattern recognition challenges.

4. Proposed Method

This unit describes the methodology and pipeline adapted to implement and assess Kannada to English neural machine translation (NMT) system. We broadly define data collection, pre-processing ideas, model architecture, training protocols, and inference mechanisms. Moreover, we deliberate the integration of transfer learning and our evaluation approaches, emphasizing how Recurrent Neural Networks (RNNs) with attention mechanisms, together with pre-trained Transformer-based models, can augment complete translation quality. The following algorithm indicates the complete translation process.

Algorithm: **Kannada-English Translation**

This algorithm takes input as Kannada text, translates it into English text and predicts desired prescription to the patient.

Step1 : Start

Step2 : Convert Patient's Speech (Kannada) to Kannada Text

Step3: Data Preparation

- a) **Text Cleaning:** Convert the input text to lowercase, exclude special characters, and strip redundant whitespace.
- b) **Tokenization:** Divide text into separable words or sub-words, then map each word to a distinctive integer.
- c) **Vocabulary Creation:** Construct mappings for both the source language and target language lexicons.

Step4: Encoding the Source Text

- a) **Word Embedding:** Translate tokenized integer sequences into dense vector form. (e.g., using Word2Vec or an Embedding layer).
- b) **Encoder:** Give each embedded source word to an Encoder LSTM. The encoder computes the whole input sequence and compresses its meaning into a fixed-dimension vector called as the "hidden state" and "cell state"

Step5: Decoding and Generating the Target text

- a) **Initial State:** Send the final hidden and cell states from the Encoder to reset the Decoder LSTM.
- b) **Start Token:** Deliver a distinct start-of-sequence token (e.g., <start>) as the initial input to the Decoder.
- c) **Teacher Forcing:** Through training, provide the tangible target translation (shifted by one step) as input to the Decoder, permitting it to predict the next word properly.
- d) **Inference:** Through real translation, the output produced by the Decoder in the preceding step is fed back into it as the next input

Step6:Output and Prediction

- Softmax Layer:** The hidden state output from the Decoder LSTM is delivered over a dense layer using a softmax activation function.
- Probability Distribution:** Softmax assesses the likelihood of each single word in your target language vocabulary. The word with the maximum probability is designated as the translation for that step.

Step7: After text extraction from the translated text model predicts the corresponding medicine from the dataset.

Step8: Stop.

5. Results and Discussions.

Our dataset encompasses over 900 samples covering patient information, with disease descriptions in Kannada together with their equivalent English translations. Using this input information, the model infers key text to summarize the disease precisely and deliver the suitable prescription. Sample data from this dataset are given in Table 1. In table 2 depicts the disease prediction results of the proposed model, emphasizing instances where Google translator and Whisper translator struggled. Table 3 summaries machine translation performance metrics, representing that our model attains higher accuracy than both benchmarks. Precisely, the proposed model verified higher BLEU and METEOR scores along with a lower Translation Edit Rate (TER) compared to Google Translator and Whisper.

Table 1: Outcome of Translation, Text Extraction and Medicine prescription

Kannada Text	English Text	Text Extraction	Predicted Medicine
ನಾನು ಕಳೆದ ನಾಲ್ಕೈದು ದಿನಗಳಿಂದ ಜ್ವರದಿಂದ ಬಳಲುತ್ತಿದ್ದೇನೆ	I have been suffering from fever for the past 4-5 days	fever+4-5+days	med1
ನನಗೆ ಕಳೆದ ಮೂರು-ನಾಲ್ಕು ದಿನಗಳಿಂದ ಜ್ವರ, ನೆಗೆಡಿ ಮತ್ತು ತಲೆನೋವು ಇದೆ	since 3-4 days I have been suffering from fever cold headache	fver+days+cold+headache	med3
ನಾನುನಿನ್ನ ರಾತ್ರಿಯಿಂದ	I have been suffering from fever since last night	night+fever	med4

ಜ್ವರದಿಂದ ಬಳಲುತ್ತಿದ್ದೇನೆ			
ನಾನು ಕಳೆದ ನಾಲ್ಕೈದು ದಿನಗಳಿಂದ ವಾಂತಿಯಿಂದ ಬಳಲುತ್ತಿದ್ದೇನೆ	I have been suffer- ing from vomiting for the past 4-5 days	vomiting+past+4- 5+days	med21
ನನಗೆ ಬೆಳಗಿನಿಂದ ತುಂಬಾ ಸುಸ್ತಾಗುತ್ತಿದೆ	I have been suffer- ing from fatigue since morning	fatigue+morning	med26
ನನಗೆ 3-4 ದಿನಗಳಿಂದ ಶೀತವಾಗಿ	since 3-4 days I have been suffer- ing from cold	cold+3-4+days	med31
ನನಗೆ 3-4 ದಿನಗಳಿಂದ ಭೇದಿಯಾಗುತ್ತಿದೆ	since 3-4 days I have been suffer- ing from diarrhea	days+diarrhea+3-4	med19

Table 2: Disease names predicted by proposed system

Disease Name	Google Translator	Whisper Translator	Proposed Model
ತಿಂಡಿ (North Karnataka)	Breakfast	Breakfast	Itching
ನವೆ (Dakshin Kannada)	Irritation	Irritation	Itching
ತುರಿಕೆ (South Karnataka)	Itching	Itching	Itching
ದಮ್ಮ	Strength	Strength	Asthma
ತುರಿ ಗಜ್ಜಿ	Scabies	Scabies	Fungal Infection
ಬಿಳಿ ಚರ್ಮ	White Skin	White Skin	Leucoderma
ನರ ಊತ	Nerve Swelling	Nerve Swelling	Varicose Veins

Table 3: Machine Translation metrics outcomes

Model	BLEU Score (↑)	METEOR R (↑)	TER (↓)
Google Translator	35.8 – 28.4	39.5%	69.2%
Whisper Translator	39.8 – 30.8	41.5%	64.6%
Proposed Model	41.5– 31.8	52.2%	53.1%

6. Limitations

The proposed model attains 95% accuracy in predicting medications, assessed against prediction errors. To ensure safety, all predicted prescriptions are swotted by medical specialists prior to supervision. As artificial intelligence models do not assurance seamless accuracy and can sporadically be prone to overfitting, human oversight remains indispensable.

7. Conclusion

In this paper we have used kannada language as natural language and implemented machine translation process using Recurrence Neural Networks (LSTM). Another objective of this paper is to identify disease with synonyms where google translator is difficult to identify the same disease with different words (synonyms). In some regional languages people uses different names for same disease, depending on regional places. Our proposed system able to identify disease name from synonyms and suggests corresponding medicine approximately. This system helps doctors to control high volume of patients in their hospital, we have implemented Kannada to English text translation using Recurrence Neural Networks. The translation comparison is made with google and whisper translator, our system gives better outcome

Competing Interests : Not Applicable

Funding Information :Not Applicable

Author contribution : Author 1 is implemented and author 2 reviewed and updated

Data Availability Statement : Created own dataset

Research Involving Human and /or Animals : No

Informed Consent : : Yes

References

1. Hugo Touvron, Louis Martin, Kevin Stone, Peter Al bert, AmjadAlmahairi, YasmineBabaei, Nikolay Bashlykov, SoumyaBatra, Prajjwal Bhargava, ShrutiBhosale, Dan Bikel, Lukas Blecher, Cristian Can ton Ferrer, Moya Chen, GuillemCucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, VedanujGoswami, Na man Goyal, Anthony Hartshorn, SagharHosseini, RuiHou, HakanInan, Marcin Karda, Viktor Kerkez, MadianKhaba, Isabel Kloumann, ArtemKorenev, Punit Singh Koura, Marie-Anne Lachaux, ThibautLavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yun ing Mao, Xavier

Martinet, TodorMihaylov, Pushkar Mishra, Igor Molybog, YixinNie, Andrew Poulton, Jeremy Reizenstein, RashiRungta, KalyanSaladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, IliyanZarov, Yuchen Zhang, An gela Fan, Melanie Kambadur, SharanNarang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

2. AkshaySwaminathan, IvánLópez, Rafael Antonio Garcia Mar, Tyler Heist, Tom McClintock, Kaitlin Caoili, Madeline Grace, Matthew Rubashkin, Michael N Boggs, Jonathan H Chen, et al. 2023. Natural language processing system for rapid detection and intervention of mental health crisis chat messages. *NPJ Digital Medicine*, 6(1):213.
3. Suresh BudhaDahal and Muhammad Aoun. 2023. Exploring the role of machine translation in improving health information access for linguistically diverse populations. *Advances in Intelligent Information Systems*, 8(2):1–13
4. Anirudh Mani, ShrutiPalaskar, NimshiVenkatMeripo, Sandeep Konam, and Florian Metze. 2020b. Asr error correction and domain adaptation using machine translation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6344–6348. IEEE.
5. Pierrette Bouillon, Glenn Flores, Maria Georgescul, Sonia Halimi, Beth Ann Hockey, Hitoshi Isahara, Kyoko Kanzaki, Yukie Nakao, Manny Rayner, Mari anneSantaholma, et al. 2008. Many-to-many multi lingual medical speech translation on a pda. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*.
6. Khai Le-Duc. 2024. Vietmed: A dataset and benchmark for automatic speech recognition of vietnamese in the medical domain. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC COLING 2024)*, pages 17365–17370.
7. Priya Ba, JM Nandhini, and Gnanasekaran Tc. 2021. An Analysis of the Applications of Natural Language Processing in Various Sectors. (2021).
8. Ignatius Ezeani, Paul Rayson, Ikechukwu Onyenwe, Chinedu Uchechukwu, and Mark Hepple. 2020. Igbo-english machine translation: An evaluation benchmark. *arXiv preprint arXiv:2004.00648* (2020).
9. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
10. Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* (2015).
11. Suraj Nair, Eugene Yang, Dawn Lawrie, Kevin Duh, Paul McNamee, Kenton Murray, James Mayfield, and Douglas W Oard. 2022. Transfer Learning Approaches for Building Cross-Language Dense Retrieval Models. *arXiv preprint arXiv:2201.08471* (2022)
12. Toan Q Nguyen and David Chiang. 2017. Transfer learning across low-resource, related languages for neural machine translation. *arXiv preprint arXiv:1708.09803* (2017)
13. Baosong Yang, Longyue Wang, Derek Wong, Lidia S Chao, and Zhaopeng Tu. 2019. Convolutional self-attention networks. *arXiv preprint arXiv:1904.03107* (2019)
14. Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. 523–530.
15. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014)
16. Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* (2015)..

17. Baosong Yang, Longyue Wang, Derek Wong, Lidia S Chao, and Zhaopeng Tu. 2019. Convolutional self-attention networks. arXiv preprint arXiv:1904.03107 (2019).
18. Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. arXiv preprint arXiv:1803.02155 (2018).
19. Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. 2016. Neural machine translation in linear time. arXiv preprint arXiv:1610.10099 (2016).
20. James Bradbury, Stephen Merity, Caiming Xiong, and Richard Socher. 2016. Quasi-recurrent neural networks. arXiv preprint arXiv:1611.01576 (2016).
21. Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional sequence to sequence learning. In International Conference on Machine Learning. PMLR, 1243–1252.
22. Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd international conference on Machine learning. 369–376.
23. Graham Neubig. 2017. Neural machine translation and sequence-to-sequence models: A tutorial. arXiv preprint arXiv:1703.01619 (2017).
24. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. Advances in neural information processing systems 30 (2017).
25. Kawin Ethayarajh. 2019. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. arXiv preprint arXiv:1909.00512 (2019).
26. Philip A Gregory, Andrew G Bert, Emily L Paterson, Simon C Barry, Anna Tsykin, Gelareh Farshid, Mathew A Vadas, Yeesim Khew-Goodall, and Gregory J Goodall. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nature cell biology 10, 5 (2008), 593–601.
27. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research 21, 1 (2020), 5485–5551.
28. Jasmijn Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. 2017. Graph convolutional encoders for syntax-aware neural machine translation. arXiv preprint arXiv:1704.04675 (2017).
29. Lilapati Waikhom and Ripon Patgiri. 2021. Graph neural networks: Methods, applications, and opportunities. arXiv preprint arXiv:2108.10733 (2021).
30. Ocheme Anthony Ekle and William Eberle. 2024. Anomaly Detection in Dynamic Graphs: A Comprehensive Survey. ACM Transactions on Knowledge Discovery from Data (2024).
31. Yangshengyan Liu, Fu Gu, Xinjian Gu, Yijie Wu, Jianfeng Guo, and Jin Zhang. 2022. Resource recommendation based on industrial knowledge graph in low-resource conditions. International Journal of Computational Intelligence Systems 15, 1 (2022), 42.
32. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems 33 (2020), 1877–1901.
33. Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Paul Barham, Xiaodong Guo, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. arXiv preprint arXiv:2010.11934 (2021).

34. Nils Reimers and Iryna Gurevych. 2020. Multilingual sentence embeddings using distillation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 4512–4525.
35. Sheng Wang, Dian Yu, Yizhong Zhang, Yejin Choi, Jianfeng Gao, et al. 2023. Text Embeddings by Weakly-Supervised Contrastive Pre-training. arXiv preprint arXiv:2302.13952 (2023)
36. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).
37. Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015).
38. Alex M Lamb, Anirudh Goyal ALIAS PARTH GOYAL, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. 2016. Professor forcing: A new algorithm for training recurrent networks. Advances in neural information processing systems 29 (2016)
39. Pradeep K. Debabala S., Prasant K. 2019. Advances in Intelligent Systems and Computing. O'Reilly Media, Inc
40. Palash Goyal, Sumit Pandey, and Karan Jain. 2018. Deep learning for natural language processing. New York: Apress (2018).
41. Alex M Lamb, Anirudh Goyal ALIAS PARTH GOYAL, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. 2016. Professor forcing: A new algorithm for training recurrent networks. Advances in neural information processing systems 29 (2016).
42. Xuan-Hien Le, Hung Viet Ho, Giha Lee, and Sungho Jung. 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11, 7 (2019), 1387