CLASSIFYING EMPATHY IN TEXTUAL ANAMNESIS USING SINGLE LAYERED LSTM PERFORMANCE

Yoenie Indrasary[1*], Metty Verasari[2], Rahadian Yusuf[3], Ary Setijadi[4], Syamsudin Noor[5], Syaifullah[6], Joni Riadi[7], and Vidya Ikawati[8]

[1,3,4,8]Institut Teknologi Bandung, Indonesia
[1,2,5,6,7]Politeknik Negeri Banjarmasin, Indonesia
[2]Universitas Mercubuana, Indonesia
[2]Division of Research of Clinical Psychology Association Region Yogyakarta
(E-mail: 33221303@std.stei.itb.ac.id)

ABSTRACT

This study explores the classification of empathy in clinical conversations using Natural Language Processing (NLP) and a single-layer Long Short-Term Memory (LSTM) model. Empathy is a critical component of patient-centered care, yet detecting and analyzing it in text-based clinical anamnesis remains underexplored. The dataset, comprising transcribed clinical interviews annotated for affective and cognitive empathy, was processed using word embeddings combining Word2Vec and GloVe. An LSTM-based classifier was developed to identify empathetic expressions in text, achieving promising results with high accuracy and AUC scores. However, limitations include a simplified focus on two empathy dimensions and challenges due to dataset size and imbalance. These findings demonstrate the potential of LSTM models in understanding empathetic communication in healthcare, with implications for enhancing doctor-patient interactions through AI-driven insights.

Key words: Empathy Classifier, Anamnesis, NLP, Machine Learning

INTRODUCTION

The study conducted in the emergence of artificial intelligence in healthcare 5.0 reveals that some potential users cite the lack of empathy for the applications and the potential to limit or even eliminate human connections as deterring factors of AI acceptance in healthcare service (1). This finding is interesting considering that the main elements supporting patient-centred healthcare services are harmonious communication and interaction between patients and their families towards health service providers, which include doctors and other health workers (2-4).

The conventional method of conveying empathy is through touch, gesture, gaze, and voice. However, social networks and online communities have opened a new paradigm for expressing empathy through natural language and text-based communication. Yet with a concept as critical to augmenting patients' positive feelings as empathy, to our knowledge, there has been no single computational study to detect empathetic support at a fine-grained level (i.e., seeking or providing empathy) from text in clinical interview context.

Anamnesis, often referred to as a clinical conversation conducted to gain patient's medical history, is the initial step in a doctor's examination process when a patient visits. It involves gathering information about the patient's medical and social background related to their health concerns. This information helps doctors to gain a comprehensive understanding of the patient's health. Hampton et al. suggest that anamnesis is one of the three primary data sources influencing a doctor's diagnosis, alongside physical exams and lab results. Bernard Lown also notes that anamnesis contributes 75% of the information needed for a diagnosis before a physical exam. In essence, history taking, in conjunction with physical exams, plays a significant role in the diagnostic process.

Furthermore, anamnesis plays a crucial role in patient care by initiating doctor-patient interaction and communication. Empathetic interactions between doctors and patients are vital because research spanning health journals like PubMed, EMBASE, and PsychINFO (1995-2011) shows that doctors' empathetic care can reduce patient anxiety, stress, and enhance clinical outcomes.

This paper is a preliminary study of classifying the presence of empathy in some textual anamnesis. It is part of the main research that tries to develop Vidya Medic: an AI-assisted intelligent healthcare system, as suggested in (5). Some research that uses AI to classify the presence or even score the level of empathy in conversation will be presented in Section II. Section III will explain the basic concept of some architecture that was proven to succeed in classifying the level of empathy in conversation. Section IV will explain the methodology of the experiment reported in this paper, and finally, the result and discussion will be presented in Section V.

LITERATUR REVIEW

Clinical Empathy

Some scholars describe empathy as 'feeling alongside' the patient (10-12), while others define it as the cognitive ability that enables healthcare professionals to comprehend the patient. Those who adopt the latter perspective create a distinction between the cognitive aspect of understanding others (referred to as cognitive empathy) and the emotional aspect of feeling for them (which they consider as sympathy), as well as feeling with them (associated with emotional empathy). Nevertheless, it is a common tendency among academics to categorize empathy into two primary types: cognitive empathy and emotional empathy.

1. Cognitive Empathy

Cognitive Empathy encompasses the recognition of others' emotions and the attribution of their mental states. In essence, it is comparable to the 'Theory of Mind' concept, as detailed by (13) and referenced in (14). This cognitive process is accomplished through the so-called 'Simulation Theory'.

As per this theory, our comprehension of the mental states of others is achieved through the simulation of their experiences. It is our capacity to adopt the perspective of others that grants us insight into their perception of the world (15-17). This concept of empathy specifically does not involve emotional involvement.

The focus on the cognitive aspect of empathy aligns with the prevailing medical ethos that prioritizes detachment, objectivity, and uniformity. As a result, this concept of empathy has gained prominence within the realm of healthcare. In actuality, it is perceived as enabling medical professionals to uphold a professional boundary with patients, formulate impartial clinical judgments, and view patients as equally deserving of care (18-19).

2. Affective Empathy

Conversely, 'affective empathy' refers to the emotional connection that arises when one encounters another person's suffering. However, the medical community and the associated medical culture tend to dismiss this type of empathy, often considering emotional empathy as reliant on uncertain or unreliable emotions (20).

In contrast to this perspective, there is a rising endorsement of the notion that affective empathy aids in understanding the patient's emotional condition and circumstances. Furthermore, emotional involvement is regarded as advantageous, as it can ignite a physician's aspiration to 'heal,' transcending the mere obligation to eliminate the ailment.

Nonetheless, because of a deep-seated concern about becoming too emotionally attached to patients, emotional involvement has been recognized as a primary contributor to the decline of impartiality and the emergence of emotional strain (21).

As previously noted, maintaining impartiality is fundamental in many clinical scenarios and medical interventions, particularly in situations involving complex diagnoses or the need for invasive treatments (20). It is, in fact, a common belief that emotions can hinder the attainment of positive medical results in sensitive situations (18-19).

LSTM Architecture (22)

We considering LSTM for sentiment analysis because past studies have demonstrated its strong accuracy in classifying conversational text sentiments. Moreover, it boasts lower resource demands and greater adaptability in generalization compared to the transformer model (23). The LSTM architecture comprises a single component known as the memory unit. Within this memory unit, four feedforward neural networks are integrated, each featuring an input and an output layer. These neural networks establish connections between input and output neurons. Consequently, the LSTM unit incorporates four fully connected layers. Among these four feedforward neural networks, three are tasked with information selection and are referred to as the forget gate, the input gate, and the output gate. These gates serve the essential functions of memory management, encompassing the removal of information from memory (forget gate), the addition of new data to memory (input gate), and the utilization of existing information in memory (output gate). The fourth neural network, known as the candidate memory, is employed to generate new data for inclusion in the memory.

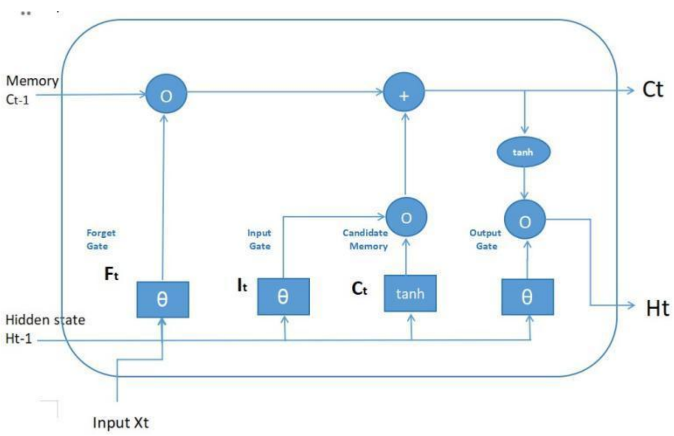


Figure 1. Architecture of a LSTM Unit

Several Works on Classifying Empathy With NLP [22]

Firoj Alam was one of the first to research the empathy aspect of conversation using NLP. Firoj Alam (6) processes audio signals from customer and agent conversations by annotating emotions in each segment of conversational audio recordings. The classifier model used is Support Vector Machines. Then Sharma et al. followed research on empathy analysis using NLP techniques in textual conversations between psychologists and clients. Here, Sharma (7) proposes an empathy analyzer model through two separate Roberta encoders to process texts originating from clients and psychologists. The output of each encoder then becomes the input of one attention block. This attention block is then connected to two linear layers, each of which is tasked with identifying empathy and extracting the rational level of the psychologist's response text (words used as indicators for assessing the three aspects of empathy). Based on the experimental results, the model proposed by Sharma et al. was able to provide the best metric values (accuracy and F1) for both empathy identification and rational extraction tasks on both types of datasets, namely TalkLife and Reddit.

Meanwhile, Vázquez EC et al. (8) conducted an empathy analysis on the empathic conversation database compiled by Rashkin et al. (9). The entire conversation is first extracted to obtain 38 conversation features with the API. Then, the results of the information extraction are classified by several classifiers to compare their performance. The classifier proposed in Vazquez's research is pattern-based classification PBC4cip. From the experimental results, it was found that the best Closeness Evaluation Measure (CEM) and Area Under the ROC Curve (AUC) metrics were given by the PBC4cip classifier.

METHOD

Objective

The objective of this study is to figure out if when doctors and patients talk to each other in clinical settings, they show empathy, and we're doing this by using NLP techniques and the LSTM model to classify their conversations.

Methodology

In the experiment, the data included audio recordings of medical students conducting clinical interviews through role-play. These audio recordings were carefully transcribed to ensure that every word matches exactly what was spoken in the audio. The transcription is then formatted to be compatible with the machine learning model for further manipulation.

In this research, the content within the transcription file, initially in .txt format, underwent a transformation, resulting in the creation of four distinct columns. The first column, designated as "doctor," encompasses discussions originating from medical professionals, while the second column, named "patient," includes dialogues originating from individuals seeking medical care. The third column, labeled "AE," was used to record a value of 1 when the conversation was evaluated as containing elements of affective empathy and 0 when it did not. Similarly, the fourth column, titled "CE," was utilized to assign a value of 1 if the conversation exhibited elements of cognitive empathy and 0 if it did not. The annotation process for determining AE and CE values was conducted by psychologists affiliated with the research division of the Clinical Psychology Association region Yogyakarta.

Following the annotation, a word embedding procedure is conducted, aiming to convert each word in a .csv dataset file into a numerical vector that represents its semantic context within a specific subject area. This operation is commonly referred to as "word embedding." In this particular research, the word embedding technique utilizes a combination of the word2vec method along with glove embedding. Glove was selected due to its demonstrated superiority in comparison to numerous conventional Word2vec models, especially when evaluated in terms of word analogy tasks. An advantage of GloVe is its capacity to directly model relationships, as opposed to acquiring them incidentally during the training of a language model (25). In this specific study, a vector dimension of 100 was opted for, encompassing 400,000 word vectors.

A portion of the word embedding result matrices from the dataset is integrated into the training phase, and the remaining part is employed for the model's validation after the training. The split between the data utilized for training and validation is set at 80% for training and 20% for validation. Notably, the model employed in this experiment is LSTM.

In a broad sense, our experimental approach can be outlined using the diagram below:

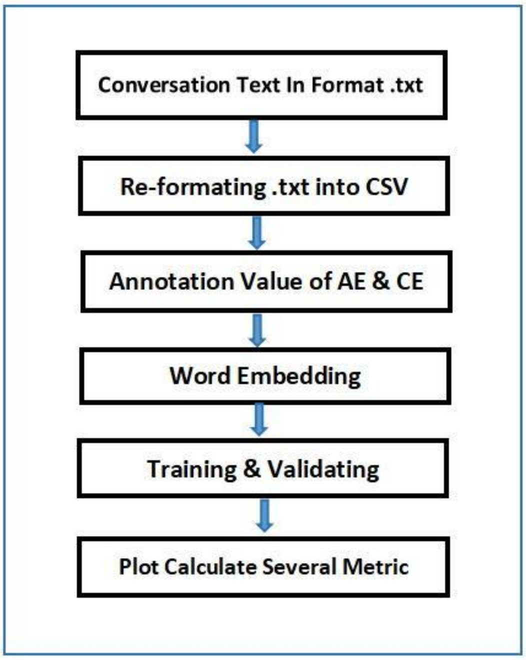


Figure 1. The Experiment Method

Dataset

The dataset used in this study is a subset of the dataset referenced in (23). It was curated by a team consisting of resident doctors specializing in internal medicine, phsyciatry, anatomical pathology, family medicine, and senior Canadian medical students. The dataset comprises medical interviews conducted in the format of Objective Structured Clinical Examinations (OSCE). These interviews cover a range of medical conditions, including respiratory, musculoskeletal, cardiac, dermatological, and gastrointestinal diseases. Notably, a significant portion of the dataset focuses on respiratory cases. For a visual overview of the case types, please refer to Figure 1. The original audio recordings were transcribed, manually reviewed for speech-to-text errors, and annotated to identify the speakers.

Table 1. Proportion of cases in the dataset's clinical dialog

No.	Cases	∑ Clinical Conversation
1	CAR	5
2	DER	1
3	GAS	6
4	GEN	1
5	MSK	9
6	RES	8
Total		**30**

Word Embedding

In order to perform the conversion that allows for a numerical representation of each token, we utilize word embedding on every token within the conversation transcription. In this research, we employ a combination of the word2vec method and glove embedding for this purpose.

Glove was chosen because it was proven to outperform many common Word2vec models on the word analogy task. One benefit of GloVe is that it is the result of direct modeling relationships, instead of getting them as a side effect of training a language model (25). In this experiment, the vector dimension we chose was 100 which consists of 400,000 word vectors.

Model Architecture

In this experiment we try to apply the approach used to carry out sentiment analysis on conversations in text. The model used as a classifier is a single layer LSTM. The architecture of LSTM that we build to become the classifier can be ilustrated as follows:

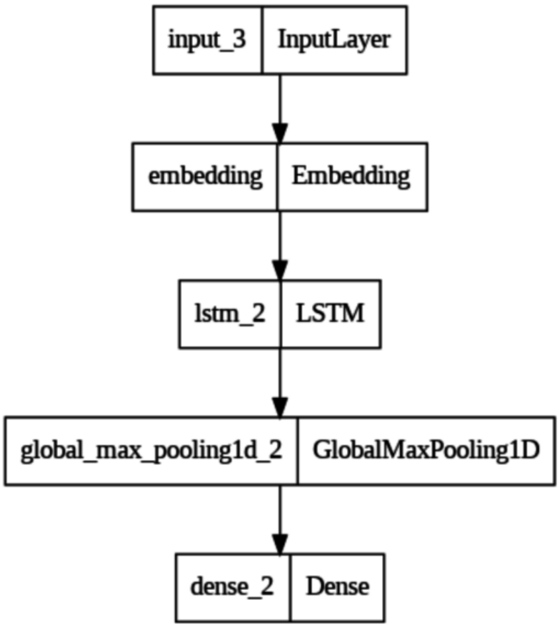
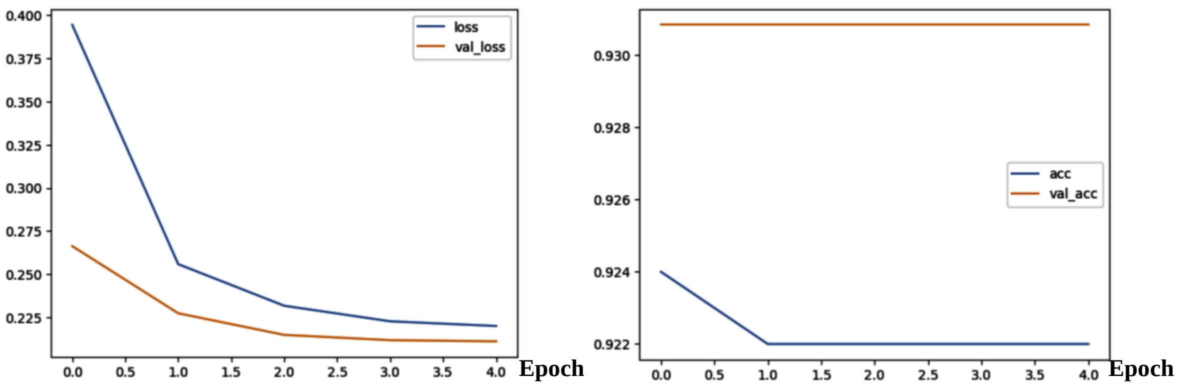


Figure 2. Architecture of LSTM used as the classifier

RESULT AND DISCUSSION

The results of the training process with batch parameters = 128, validation split = 0.2, epoch = 5, learning rate = 0.01, loss function binary cross entropy, and optimizer Adam are as shown in the graph as follows:

Loss-function **Accuracy**



(a) **(b)**

Figure 3. (a) Loss-Function value with epoch and (b) Accuracy of value with epoch of one layer lstm classifier with 15 memory units

Meanwhile, the average AuC value is equal to 1.

This experiment's findings suggest that the LSM model has been developed for the purpose of classifying empathy levels in clinical conversations. However, there are limitations to this study, including the simplification of empathy aspects, which focus solely on affective and cognitive empathy. Additionally, the model validation process should consider more reliable methods like k-fold cross-validation. The experiment's constraints, such as the limited number of datasets and their unbalanced characteristics, hinder the representation of semantic features in more complex conversations.

CONCLUSION

According to the findings from our experiments mentioned earlier, we can deduce that utilizing the LSTM architecture yields promising outcomes in determining whether doctors' communication in clinical interviews contains empathetic elements or not.

Acknowledgment: The authors would like to express their appreciation for the support of Division of Research of Clinical Psychology Association Region Yogyakarta for the annotation process.

REFERENCES

1. J. Al-Jaroodi, N. Mohamed, N. Kesserwan and I. Jawhar, "Human Factors Affecting the Adoption of Healthcare 4.0," 2023 IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 2023, pp. 1-7, doi: 10.1109/SysCon53073.2023.10131064.
2. Bhargava K, Bhargava D. Evidence Based Health Care: A scientific approach to health care. Sultan Qaboos Univ Med J. 2007 Aug;7(2):105-7. PMID: 21748091; PMCID: PMC3074860.
3. G.L.Tortorella, F.S. Fogliatto, K.F. Espôsto. Effects of contingencies on healthcare 4.0 technologies adoption and barriers in emerging economies. Technol Forecast Soc Change, 156 (2020), p. 120048.
4. G.L. Tortorella, T.A. Saurin, F.S. Fogliatto, V.M. Rosa, L.M. Tonetto, F. Magrabi Impacts of Healthcare 4.0 digital technologies on the resilience of hospitals. Technol Forecast Soc Change, 166 (2021), p. 120666.
5. Vidya Ikawati Ary Setijadi. "VidyaMedic: Smart Alert System For Continues Vital Sign Monitoring", 2023 IEEE International Conference on Electrical Engineering and Informatic, Bandung, Indonesia.
6. F. Alam, M. Danieli and G. Riccardi, "Can we detect speakers' empathy?: A real-life case study," 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, Poland, 2016, pp. 000059-000064, doi: 10.1109/CogInfoCom.2016.7804525Rew
7. Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 5263–5276, Online. Association for Computational Linguistics
8. Montiel-Vázquez EC, Ramírez Uresti JA, Loyola-González O. An Explainable Artificial Intelligence Approach for Detecting Empathy in Textual Communication. Applied Sciences. 2022; 12(19):9407. https://doi.org/10.3390/app12199407
9. Rashkin, H.; Smith, E.M.; Li, M.; Boureau, Y.L. Towards empathetic open-domain conversation models: A new benchmark and dataset. In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019. [Google Scholar]Et
10. Trevithick, Pamela. 2005. Social work skills. Berkshire, UK: Open University Press.Te
11. Boulton, Robert. 1987. People skills. Sydney, New South Wales, Australia: Simon & Schuster.
12. Gair, Susan. 2012. Feeling their stories: Contemplating empathy, insider/outsider positionings, and enriching qualitative research. Qualitative Health Research 22 (1): 134–214.
13. Schulte-Rüther, Martin, Hans J. Markowitsch, Gereon R. Fink, and Martina Piefke. 2007. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. Journal of Cognitive Neuroscience 19 (8): 1354–1372.Et
14. Finset, Arnstein. 2010. Conceptual explorations on person-centered medicine 2010: emotions, narratives and empathy in clinical communication. International Journal of Integrated Care. https://doi.org/10.5334/ijic.490.Wte
15. Gladstein, Gerald A. 1983. Understanding empathy: Integrating counseling, developmental, and social psychology perspectives. Journal of Counseling Psychology 30 (4): 467.
16. Rushton, J. Philippe. 1980. Altruism, socialization, and society. Prentice-Hall.
17. Zeyer, P. Bahar, K. Irie, R. Schlüter and H. Ney, "A Comparison of Transformer and LSTM Encoder Decoder Models for ASR," 2019 IEEE Automatic Speech Recognition and Understanding Workshop

(ASRU), Singapore, 2019, pp. 8-15, doi: 10.1109/ASRU46091.2019.9004025.

18. Pedersen, Reidar. 2008. Empathy: A wolf in sheep's clothing? Medicine, Health Care and Philosophy 11 (3): 325.
19. Coulehan, John L. 1995. Tenderness and steadiness: Emotions in medical practice. Literature and Medicine 14 (2): 222–236.
20. Halpern, Jodi. 2011. 13 Clinical empathy in medical care. In Empathy: From bench to bedside, 229–244. Cambridge: MIT Press.
21. Halpern, Jodi. 2014. From idealized clinical empathy to empathic communication in medical care. Medicine, Health Care and Philosophy 17 (2): 301–311.
22. Halpern, Jodi. 2001. From detached concern to empathy: Humanizing medical practice. Oxford: Oxford University Press.
23. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks. ArXiv. /abs/1909.09586
24. Yang, Shuyuan, et al. "A New Directional Multi-resolution Ridgelet Network." Frontiers of Electrical and Electronic Engineering in China, 2005, https://doi.org/10.1007/s11460-008-0026-2.
25. Fareez, F., Parikh, T., Wavell, C. et al. A dataset of simulated patient-physician medical interviews with a focus on respiratory cases. Sci Data 9, 313 (2022). https://doi.org/10.1038/s41597-022-01423-1
26. Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.