International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Development of Natural Language Dialogue System for Indian Language in Healthcare Domain

Saritha Shetty[1], Sarika Hegde[2], Savitha Shetty[3]*

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: The healthcare dialogue system is developed using doctor-patient discussions. The doctor-patient conversations are used as input, and for every symptom, allopathic prescriptions are produced as the output. A total of 648 doctor-patient conversations were gathered for the dataset by going in-person to two rural hospitals. The use of local language for data collecting in the healthcare field is novelty in this study. Features are extracted from the words for further analysis. The linguistic model is constructed by feeding these features into a deep learning model. Using deep learning, the clinical dialogue system is developed. Evaluation result shows F1 score of 85.48%, recall of 85.31%, and precision of 85.65%. Following assessment, the ROC AUC value is 0.9233.

Keywords: *Dialogue system, Deep learning model, Healthcare, Natural Language Processing.*

1. Introduction

Allopathic medicine covers a wide range of illnesses and treatment modalities targeted at addressing various health conditions [1]. It is the predominant paradigm in many regions for healthcare. In rural healthcare settings, when access to specialized medical resources may be restricted, it becomes even more important to identify and manage common disorders. It was possible to obtain important insights into the occurrence and treatment of various allopathic disorders by the careful documentation and evaluation of doctor-patient talks. The range of allopathic diseases found in rural areas, from musculoskeletal issues to gastrointestinal disorders and respiratory infections, highlights the significance of tailored therapies and integrated healthcare systems [2].

The intricate process of identifying features from a set of doctor-patient conversations rigorously gathered through in-person research at two local allopathic hospitals is covered in the introduction. A rigorous dedication to documenting clinical presentations and associated allopathic therapy set apart the data gathering approaches used. Researchers used classic pen and paper methods to methodically describe the nuances of these interactions, which laid the groundwork for more investigation. Following the rigorous phase of data gathering, the data underwent a significant digital transformation enabled by the https://kannada.indiatyping.com/ site. This digital conversion process was a significant turning point in the research as it allowed the transition from written records to a more easily accessible and analysed format.

[1] NMAMIT (Nitte deemed to be University), Nitte-574110, INDIA
ORCID ID : 0000-0001-5575-5134
[2] NMAMIT (Nitte deemed to be University), Nitte-574110, INDIA
ORCID ID : 0000-0002-8781-9451
[3] NMAMIT (Nitte deemed to be University), Nitte-574110, INDIA
ORCID ID : 0000-0003-1295-2562
* Corresponding Author Email: shettysavi1@nitte.edu.in

1.1. Research objectives

The research indicates that little attempts have been made to develop a conversation system for healthcare, particularly for languages with less resources. It is now more crucial than ever to find medical remedies without physically visiting a doctor, both during and after the COVID-19 pandemic. Thus, a conversation system in the dialect that recommends allopathic treatment for common medical ailments is developed.

Following are the noteworthy contributions of the research.

- Collection of 648 doctor patient conversation samples in healthcare domain in local language.
- Extracting features from the data collected samples
- Building language model using deep learning
- Response generation for the input symptom in local language.

2. Literature Survey

The Code-Mixed Medicinal Task-Oriented dialogue Dataset, which includes 3005 Telugu-English Code-Mixed dialogues with 29,000 utterances across ten specialisations, was created to aid in the study and development of medical dialogue systems in multilingual settings [3]. By emphasising the methodical structuring of doctor-patient interactions, conversation analysis (CA) separates medical interviews beyond ethnographic and quantitative research by analyzing them as naturally occurring exchanges [4]. In order to address the bodily symptoms, fear, and impairments in everyday living that are not as prominent in the anxiety assessment instruments currently in use, the study created 615 Kannada items pertaining to anxiety that were divided into 21 domains. This qualitative method represents a first step towards building the PANIQ instrument, combining the personal experiences of stakeholders and anxiety sufferers [5]. Thirteen focus groups and 78 semi-structured interviews were conducted as part of this qualitative study in both urban and rural Karnataka, South India, to investigate obstacles to point-of-care testing. It was discovered that patients encounter numerous obstacles when trying to obtain diagnostic services, such as expense, the distance, social factors, and difficult interactions with medical professionals. These obstacles frequently cause delays or the patient to decide not to undergo testing; for this reason, making patients navigate through diagnostic services easier is essential to the effective rollout of new point-of-care tests [6].

Cataract Bot, an automaton driven by huge language models, was created to deliver trustworthy, expert-verified answers regarding cataract surgery in response to the deluge of conflicting and frequently erroneous health information that is readily available online. In a trial involving 49 participants, this chatbot—which was developed in partnership with a major eye hospital in India—proved its worth by offering multifunctional and multilingual support, dependable information that is easy to access, and accommodations for a range of literacy levels[7].

3. Methodology

During in-person consultations between the patient and the physician, data on doctor-patient conversations was gathered from the regional allopathic hospital. A total of 648 conversations between a physician and patient about 84 ailments were gathered. Out of 648 conversations, 48 are ignored since there is not enough information regarding the illness and recommended treatment. The information gathered is not structured. It includes details about the patient's present symptoms as well as the severity and duration of the ailment. Medications and health advice for the patient are included in this example doctor-patient conversation. For the information to be processed further, these pen and paper collected details need to be converted from an unorganized form to a structured form. To put the data in an organized fashion, information about diseases and prescription drugs are needed. Preparation and feature generation from input are shown in Fig 1.

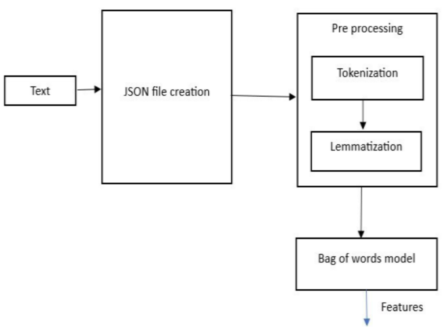

Fig. 1. Methodology utilized.

The input questions that have been gathered from the hospital are pre-processed using the processes listed below. Eliminating punctuation and other unique characters from the raw text is known as text cleaning. Characters that aren't alphanumeric and could give the model false information are removed from the input. Words with extra spaces between them are removed. This considers the provided text from the doctor-patient conversation. The conversation samples gathered from the allopathic clinics are used to construct the JSON file. For additional investigation, the JSON file's patterns have been pre-processed.

3.1. Steps in input processing

Step 1: Input text

Patients can enter single words or sentences that indicate the illness in their inquiry. The doctor has been informed of particular symptoms in these queries. For every illness, the doctor prescribes both medicine and advice. Pre-processing includes tokenization and lemmatization.

Step 2: The construction of the JSON file

JSON or JavaScript Object Notation, is an a text-based convention for encoding structured data. The questions and answers the patient provides to the doctor are used to construct the JSON file. The doctor's questions to the patient become pattern in the file's JSON format as depicted in Fig 2. The doctor's recommended prescription becomes the action. The doctor's recommendations may also be included in the response. The input phrase loses all symbols and special characters.

```
{"intents" : [
    {"tag" : "knee ache",
        "patterns" : ["I have Knee ache doctor", "i have Knee ache","Knee ache"],
        "responses" : ["apply diclo gel ointment and take rest"],
        "context" : [""] }]}
```

Fig. 2. JSON file format.

Step 3: Extraction of Features

The JSON file's contents must be pre-processed. Lemmatization, tokenization and stop word removal are aspects of pre-processing. The act of tokenizing text involves separating it into discrete tokens, which can be words, sentences, symbols, or other significant components. Words are reduced to their most basic or root form through lemmatization.

3.1.1. Tokenization

The technique of separating a text into its individual words or phrases is known as tokenization. Kannada sentences are broken down into separate words or phrases during this process. Sentence tokenization separates the content into sentences, whereas phrase tokenization separates the text into individual words.

3.1.2. Stop word elimination

Stop words are phrases that are frequently employed but don't possess a clear meaning or contribute to the overall understanding of the text. The text has stop words that are frequently used and are referred as "stop words." First, a list of stop words is created. While traversing through the tokens, words that match a stop phrase on the stop word listing are removed.

3.1.3. Lemmatization

The lemma, or dictionary form of the word, is returned by the lemmatizer after the inflectional word has had all affixes removed. The lemmatizer requires more language knowledge than the stemmer because the split lemma is a valid root and linguistically relevant phrase. The process of extracting an inflectional word's stem by eliminating its affixes is known as stemming. The root word's meaning is lost during stemming. Lemmatization is therefore recommended for pre-processing.

Step 4: Bag of words model

A bag of words (BoW) model is the encoding technique used to encode text data. It is a method for taking textual information and converting them into numerical forms that can be used by machine learning techniques. The term's presence or absence from the set of phrases is indicated by the Binary Bag of Words. The element in the vectors (1 or 0) indicates it. For example, the feature vector's number 1 is inserted if the phrase "headache" appears in the manuscript. Words can also be transformed into vectors by using word embeddings.

4. Deep learning model architecture

Deep learning model parameters are depicted in **Table 1**. **Figure 3** depicts the layers of deep learning model. A three-layered deep artificial neural system is constructed in Keras. In order to anticipate the output intents using the soft max, 128 neurons make up the first layer, 64 neurons make up the second layer, and precisely a comparable number of neurons make up the third output layer. Because it can speed up the gradient and produce reliable model outputs, stochastic-gradient-descent with Nesterov is used as an optimizer. Training can begin by adapting a prototype to the training set. The mathematical model is saved when it is trained. The model created will be maintained as an instance of pickle so that it may be imported at a later time, preventing the need to repeat the training procedure. The provided words are processed by a bag-of-words method to extract characteristics.

The design of an artificial neural network intended for dialogue system development is depicted in the flowchart that is provided.

Input Layer : Data enters the neural network through this layer. When a dialogue system is used, its input data is sent to it as a word bag containing user queries.

Dense layer (128): Using the rectified line unit (ReLU) activating function, it applies linear modifications to the input data and provides non-linearity.

Dropout (0.5): To avoid overfitting, dropout layers are added for regularization. This layer forces the network to acquire more resilient characteristics by randomly dropping 50% of its neurons during training (dropout rate of 0.5).

Dense layer (64): In this layer, higher-level features are extracted from the data by additional processing.

Dropout (0.5): By arbitrarily deactivating neurons while training, this dropout layer, which has a dropout rate of 0.5, aids in regularization by lessening the network's dependence on particular features.

Dense layer (Number of Classes): This layer is subjected to the SoftMax activation function, which turns the raw outputs score into probabilities and permits multi-class categorization.

Table 1. Deep learning model parameters.

Model parameters	
Learning Rate	0.01
Batch Size	128
Optimizer	Stochastic gradient descent
Loss function	Binary cross-entropy
Dense layer	1

5. Results and Discussion

Training accuracy is a measure of how well the deep learning model performed during the training phase. Training accuracy for a learning rate of 0.01 is shown in the Fig. 4. The accuracy graph got better when the epoch count enhanced. Training loss is depicted in Fig 5.

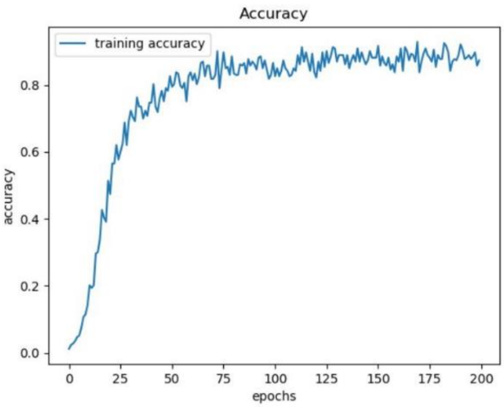

Fig. 4. Training accuracy

The lemmatized phrases that were created from the supplied patterns serve as the data source to the bag of phrases architecture. During implementation, the binary bag of phrases is employed. It returns 0 or 1 depending on whether a specific term is present in the given context or not.

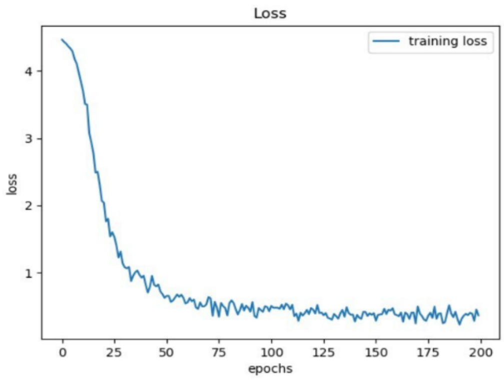

Fig. 5. Training loss

Figure 6 displays a bag of words made for 127 distinct lemmatized words. These lemmatized words are fed into the deep learning algorithm as numerical vector input features.

```
p: OUPUT OF BOW MODEL [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

Fig. 6. Bag of words output

6. Model Evaluation

The percentage of accurate classifications can also be used to gauge binary classification accuracy, as depicted in (1).

$$Accuracy = \frac{(\mathrm{TN+TP})}{(\mathrm{FN+FP+TP+TN})} \quad (1)$$

Precision seeks to determine the proportion of affirmative identifications that were actually accurate. The definition of precision is as depicted in (2).

$$Precision = \frac{\mathrm{TP}}{(FP+TP)} \quad (2)$$

The recall is denoted by (3).

$$Recall = \frac{\mathrm{TP}}{(FN+TP)}$$

(3)

The recall as well as accuracy scores' mathematical mean is denoted by the F1 score in (4).

$$\mathrm{F1-score} = 2 * \frac{(recall*precision)}{(recall+precision)} \quad (4)$$

The results revealed an F1 score of 85.48, recall of 85.31, a nd precision of 85.65. The Fig 7 showcases the results after evaluation. Figure 8 depicts the precision-recall curve for the dialogue system.

A higher area under the curve indicates better performance. Table 2 depicts comparison.

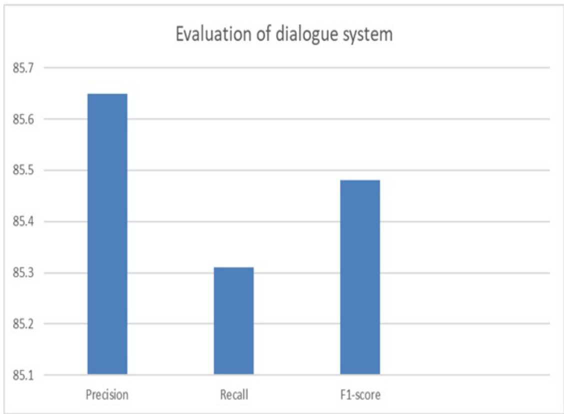

Fig. 7. Evaluation results

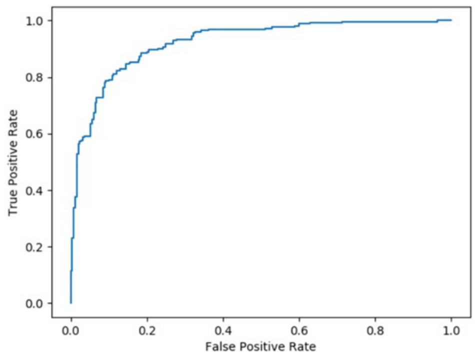

Fig. 8. Precision recall curve

Table 2. Comparison of results.

Author	Method	Accuracy
Abilkaiyrkyzy et al., 2024 [8]	Pre trained BERT models	69%
Rani & Jain, 2024 [9]	Survey paper	--

Varshney et al., 2024 [11]	Neural generative model	Classification accuracy for the Topical Chat as well as CMU_DoG collections was 0.78 and 0.73, respectively.
Samak et al., 2022 [10]	Naïve bayes classifier	-
Ramjee et al., 2024 [7]	Large language models	-

5. Conclusion

Research introduces a novel approach to developing a healthcare dialogue system utilizing doctor-patient conversations. By collecting 648 real-world interactions from rural hospitals, the research highlights the importance of incorporating regional languages into healthcare technology. Features extracted from these dialogues were used to construct a linguistic model, which was then integrated into a deep learning framework. The resulting system demonstrated robust performance, achieving an F1 score of 85.48%, recall of 85.31%, and precision of 85.65%, with an impressive ROC AUC value of 0.9233. These outcomes underscore the effectiveness of the proposed deep learning model in accurately generating allopathic prescriptions from symptom descriptions, paving the way for enhanced healthcare delivery in rural and linguistically diverse regions. The precision could be enhanced by expanding the dimension of the dataset.

Acknowledgements

We sincerely thank our college for providing necessary resources.

Conflicts of interest

There exist no conflicts of interest, according to the authors.

References

[1] Ramakrishnan, P., Dias, A., Rane, A., Shukla, A., Lakshmi, S., Ansari, B. K. M., ... & Koenig, H. G. (2014). Perspectives of Indian traditional and allopathic professionals on religion/spirituality and its role in medicine: basis for developing an integrative medicine program. *Journal of Religion and Health*, *53*, 1161-1175.

[2] Xu, Y. (2024). Boundaries and classification: the cultural logic of treating foreign medicine. *Humanities and Social Sciences Communications*, *11*(1), 1-8.

[3] Dowlagar, S., & Mamidi, R. (2023). A code-mixed task-oriented dialog dataset for medical domain. *Computer Speech & Language*, *78*, 101449.

[4] Maynard, D. W., & Heritage, J. (2005). Conversation analysis, doctor–patient interaction and medical communication. *Medical education*, *39*(4), 428-435.

[5] Sowmya, H. R., Ghani, S., Sreedaran, P., Sahu, M. K., Mysore, A., & Sharan, P. (2023). Cultural Expression of Anxiety Symptoms in Kannada Language: A Qualitative Study. *Indian Journal of Psychological Medicine*, *45*(5), 496-502.

[6] Yellapa, V., Devadasan, N., Krumeich, A., Pant Pai, N., Vadnais, C., Pai, M., & Engel, N. (2017). How patients navigate the diagnostic ecosystem in a fragmented health system: a qualitative study from India. *Global Health Action*, *10*(1), 1350452.

[7] Ramjee, P., Sachdeva, B., Golechha, S., Kulkarni, S., Fulari, G., Murali, K., & Jain, M. (2024). CataractBot: An LLM-Powered Expert-in-the-Loop Chatbot for Cataract Patients. *arXiv preprint arXiv:2402.04620*.

[8] Abilkaiyrkyzy, A., Laamarti, F., Hamdi, M., & El Saddik, A. (2024). Dialogue System for Early Mental Illness Detection: Towards a Digital Twin Solution. *IEEE Access*.

[9] Rani, S., & Jain, A. (2024). Optimizing healthcare system by amalgamation of text processing and deep learning: a systematic review. *Multimedia Tools and Applications*, *83*(1), 279-303.

[10] Samak, A. R., Bafna, P. B., & Saini, J. R. (2022). A Novel Marathi Speech-Based Question and Answer Chatbot for the Educational Domain. In *ICT with Intelligent Applications: Proceedings of ICTIS 2022, Volume 1* (pp. 379-388). Singapore: Springer Nature Singapore.

[11] Varshney, J. P., & Naresh, R. (2005). Comparative efficacy of homeopathic and allopathic systems of medicine in the management of clinical mastitis of Indian dairy cows. *Homeopathy*, *94*(02), 81-85.