Chatbots in Healthcare Communication

Kunal Inglunkar (G44511973)
Data Science Department
kunal_inglunkar@gwu.edu

Xue Ming Wang (G20580112)
Statistics Department
xmwang8@gwu.edu

***Abstract*—Chatbots, an evolving facet of modern communication, have emerged as transformative tools within diverse disciplines, notably in healthcare and medicine. The persistent challenge of optimizing communication between healthcare providers and individuals necessitates innovative solutions. Integration of chatbots in healthcare systems presents a promising avenue to surmount these hurdles, aiming to enhance accessibility, accuracy, and personalization of medical assistance. However, existing platforms often lack dynamic intelligence, hindering their ability to promptly and effectively cater to diverse patient needs. This research focuses on addressing this crucial gap by pioneering a dynamic and intelligent chatbot platform leveraging deep learning techniques like word embedding and LSTM models. By harnessing the intricate capacities of artificial intelligence, this project aspires to redefine healthcare communication, bridging accessibility gaps and elevating the quality of medical assistance available to individuals. The significance of this endeavor lies in its potential to revolutionize the healthcare landscape, making immediate, accurate, and personalized medical guidance readily available to all individuals seeking assistance. This exploration into the realm of chatbots within healthcare ecosystems is poised to contribute to a paradigm shift in patient care, establishing a new standard of accessible and tailored medical support.**

Keywords—chatbot, artificial intelligence, deep learning, LSTM, word embedding, healthcare communication, medical assistance, patient care

I. Introduction

In contemporary healthcare, the optimization of communication between individuals and healthcare providers remains a persistent challenge. Existing methods often encounter hurdles in delivering immediate, accurate, and personalized medical assistance. The integration of chatbots within healthcare systems has emerged as a potential solution, aiming to streamline communication and enhance patient care. However, existing chatbot platforms often lack the dynamic intelligence necessary to cater to diverse patient needs promptly and effectively. Our project addresses this gap by focusing on the development of a dynamic and intelligent chatbot platform that leverages deep learning techniques, such as word embedding and LSTM models, to provide accurate and personalized medical assistance in real-time. The significance of this project lies in its potential to transform healthcare communication, improving patient care accessibility, and elevating the quality of medical assistance available to individuals in need. By harnessing the power of AI-driven chatbots, this research aims to revolutionize healthcare systems, making immediate and reliable medical assistance readily accessible to everyone.

II. Related work

Chatbots have emerged as pivotal tools within healthcare, revolutionizing patient interaction and information dissemination [1][3]. This literature review synthesizes findings from diverse studies exploring the development, effectiveness, and acceptance of chatbots within the healthcare domain [4][5][11].

Many studies emphasize the integral role chatbots play in facilitating seamless interactions between users and healthcare providers, aiming to streamline communication within healthcare [10]. Its objective of developing an intelligent healthcare chatbot utilizing deep learning techniques resonates with subsequent studies exploring AI-driven advancements [3][9][12].

In "Large Language Model-Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures," (Decker et al., 2023) the efficacy of a "LLM-based chatbot versus surgeons in generating readable, accurate, and comprehensive information" for informed consent demonstrates the potential of chatbots to provide understandable and detailed medical information [1]. The study's conclusion regarding the chatbot's promise in improving patient comprehension aligns with the broader goal of enhancing healthcare communication.

Moreover, "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study" (Nadarzynski et al., 2019) delves into "the acceptability and hesitancy surrounding AI-led chatbots in healthcare" [7]. This mixed-methods approach revealed user reservations regarding accuracy, security, and trustworthiness while also highlighting the potential for future acceptance. The study's emphasis on user-centered design aligns with the aim of creating effective and trustworthy healthcare chatbots [6].

Furthermore, "Chatbot for healthcare system using Artificial Intelligence" (Maher et al., 2022) and "Machine Learning in Healthcare Communication" (Siddique et al., 2021) both underline the critical role of AI and machine learning techniques, particularly in healthcare [6] [10] . Their exploration of AI-driven chatbots in providing medical information and optimizing patient care resonates with the overarching theme of leveraging technology for enhanced healthcare accessibility and efficiency.

Additionally, the systematic review by Xu et al. (2021) presents a comprehensive view of chatbot integration in cancer care [12]. It stresses the potential of deep learning methods, particularly in cancer diagnostics, aligning with the broader aim of leveraging AI to improve disease diagnosis and patient care [8][9].

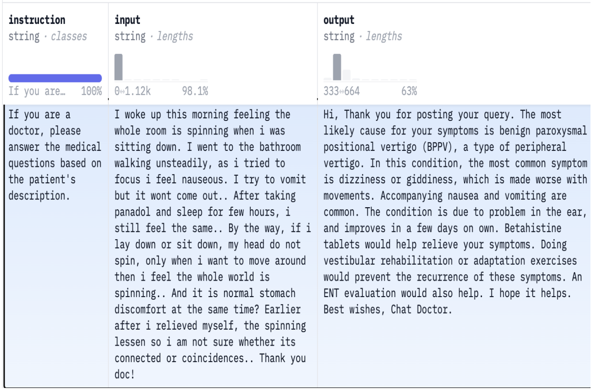
Lastly, "Analyzing the performance differences between pattern matching and compressed pattern matching on texts" (Erdogan et al, 2013) offers insights into optimizing chatbot performance by "matching patterns directly on compressed data" [2]. Its focus on enhancing efficiency aligns with the broader objective of streamlining chatbot functionalities within healthcare systems [7].

In summary, these studies collectively underscore the pivotal role of chatbots, powered by AI and deep learning, in revolutionizing healthcare communication, patient care, and information dissemination [4] [9]. They highlight the need for user-centered design, trustworthiness, and efficiency in developing healthcare chatbots, thereby paving the way for future advancements and applications in the field.

The compilation of studies presented in the related work accentuates the significant role of chatbots in modernizing healthcare communication and patient care. These analyses demonstrate the evolution of AI-driven chatbots in addressing critical challenges within the healthcare domain. The primary focus of these studies aligns with our project's vision of creating an advanced healthcare chatbot utilizing deep learning techniques. Each study highlights the importance of user-centered design, trustworthiness, efficiency, and AI integration, contributing to the foundational concepts driving our research. Collectively, these findings fortify our project's commitment to revolutionizing healthcare communication, accessibility, and the delivery of immediate and accurate medical assistance to individuals.

III. Dataset

The dataset is obtained from Hugging Face ("lavita/ChatDoctor-HealthCareMagic-100k") containing 112,165 rows and 3 columns—Instruction, Input, and Output. It comprises conversation logs that simulate interactions between individuals seeking medical advice and a healthcare chatbot system developed to respond to their inquiries or concerns. For instance, within this dataset, one dialogue involves a user detailing symptoms like dizziness, nausea, and discomfort while moving. The subsequent response generated by the chatbot suggests a potential diagnosis and offers advice on medication and exercises to alleviate these symptoms. These diverse conversations encapsulate various health-related scenarios, serving as a valuable resource for training and evaluating chatbot models within the healthcare domain. These models aim to provide medical guidance, support, or information to users based on their symptoms or health queries, showcasing the potential of AI-driven systems in assisting healthcare interactions.



instruction string · classes If you are… 100%	input string · lengths 0↔1.12k 98.1%	output string · lengths 333↔664 63%
If you are a doctor, please answer the medical questions based on the patient's description.	I woke up this morning feeling the whole room is spinning when i was sitting down. I went to the bathroom walking unsteadily, as i tried to focus i feel nauseous. I try to vomit but it wont come out.. After taking panadol and sleep for few hours, i still feel the same.. By the way, if i lay down or sit down, my head do not spin, only when i want to move around then i feel the whole world is spinning.. And it is normal stomach discomfort at the same time? Earlier after i relieved myself, the spinning lessen so i am not sure whether its connected or coincidences.. Thank you doc!	Hi, Thank you for posting your query. The most likely cause for your symptoms is benign paroxysmal positional vertigo (BPPV), a type of peripheral vertigo. In this condition, the most common symptom is dizziness or giddiness, which is made worse with movements. Accompanying nausea and vomiting are common. The condition is due to problem in the ear, and improves in a few days on own. Betahistine tablets would help relieve your symptoms. Doing vestibular rehabilitation or adaptation exercises would prevent the recurrence of these symptoms. An ENT evaluation would also help. I hope it helps. Best wishes, Chat Doctor.

Fig. 1. The content of the dataset.

Input column consists of the queries of the patients and the output column provides the response to the user queries. The instructions column is of no use it has the same text in all the rows.

IV. Proposed approaches

A. Preprocessing

Drop the 'instruction' column:

- Identify and locate the 'instruction' column in the dataset.
- Use appropriate functions or methods to drop this column.

Text Cleaning:

- Iterate through each row in the remaining text column.
- Remove unnecessary characters such as special symbols and numbers.
- Convert the text to lowercase to ensure uniformity.
- Remove punctuation from the text.

Lemmatization:

- Utilize the NLTK library for lemmatization.
- Tokenize the cleaned text into individual words.
- Apply lemmatization to obtain the base or dictionary form of each word.
- Reassemble the lemmatized words into sentences.

Stop Word Removal:

- Use NLTK to download and access a list of stop words.
- Tokenize the lemmatized text into individual words.
- Remove stop words from the tokenized text, excluding essential words that might have been stemmed.

Stemming:

- Utilize the NLTK library to perform stemming.
- Tokenize the text into individual words.
- Apply the Porter Stemmer to obtain the root or base form of each word.
- Reassemble the stemmed words into sentences.
- Save the preprocessed text data, which now lacks the 'instruction' column, in a new dataset or replace the existing one.
- The final output should consist of clean, lemmatized, stop-word-free, and stemmed text data.

B. Word Embedding

Integration of spaCy for Language Model Loading: The research employs spaCy, a popular natural language processing library, to load an English language model. SpaCy is chosen for its robustness and efficiency in various NLP tasks.

Tokenization of Text: The text undergoes tokenization, a crucial pre-processing step in NLP, using spaCy's tokenization capabilities. This process breaks down the text into individual tokens, such as words or punctuation marks, facilitating subsequent analysis.

Leveraging spaCy's Word Vectors: To enhance the representation of textual information, the study utilizes spaCy's word vectors. These vectors, generated by advanced machine learning models, capture the semantic meaning of words. This step contributes to the creation of word embeddings, which are essential for tasks like similarity analysis and context understanding.

Word Embeddings for Enhanced Representations: The tokenized text is transformed into word embeddings, a numerical representation of words, through spaCy's word vectors. This enables the model to understand the relationships and meanings between words, providing a more nuanced understanding of the text. After word embedding the output dataframe was saved to a csv file.

C. Padding

To ensure consistency in input and output dimensions, we employ padding techniques. This practice not only promotes smooth data processing but also contributes to the stability and performance optimization of our model. Specifically, we implement sequence padding, wherein both input and output sequences are padded to a standardized length (max_len). This standardization ensures uniformity in model input, fostering a seamless and efficient computational process.

Padding Sequences: The application of sequence padding involves extending both input and output sequences to conform to a predefined length (max_len), thereby establishing a standardized format for model input. This approach is instrumental in enhancing the overall robustness and performance of the model during its operation.

V. System design and implementation

Building Model

The processing of an extensive dataset led to recurrent kernel restarts with each model iteration. In response to this challenge, we initiated our analysis with a subset comprising 1000 rows as a strategic measure to alleviate kernel-related issues. To address and incrementally overcome these challenges, we systematically expanded the subset size, aiming to enhance overall system stability and ensure consistent performance throughout our experimentation process. This progressive approach allowed for a more controlled and stable exploration of the dataset, mitigating kernel disruptions and facilitating a more robust analysis of the models under consideration.

To facilitate robust model training and evaluation, the dataset is partitioned into distinct training and validation sets. This division adheres to the conventional 80-20 split, where 80% of the data is allocated for training purposes, and the remaining 20% is reserved for validation.

The architectural components of our model encompass an Embedding layer, designed to transform input sequences into dense vectors, providing a meaningful representation for subsequent processing. This is followed by an LSTM layer, which leverages long short-term memory mechanisms to capture and learn intricate patterns within the sequential data.

The subsequent inclusion of TimeDistributed Dense layers extends the model's capability by applying a dense layer independently to each time step of the sequence. This enables the extraction and utilization of temporal information at each point in the sequence, enhancing the model's ability to discern intricate patterns and dependencies.

For model compilation, we employ the Adam optimizer, a widely utilized optimization algorithm known for its efficiency and effectiveness in training neural networks. The loss function chosen for this architecture is sparse categorical cross-entropy, aligning with the nature of the classification task. This combination of optimizer and loss function is tailored to maximize the model's performance in accurately categorizing sequences. The compilation process lays the foundation for subsequent training, ensuring the model is primed to effectively learn and generalize from the provided sequential data.

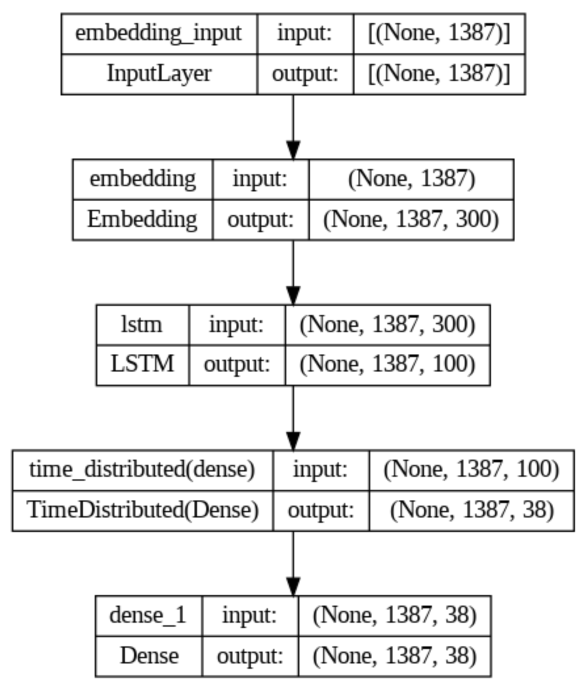


Fig. 2. The architecture of our model.

For model training, we trained the model for 10 epochs with batch size of 64. After train the model we got accuracy of 0.88 on the train dataset. On evaluating the model on test data we got an accuracy of 0.88.

Accuracy

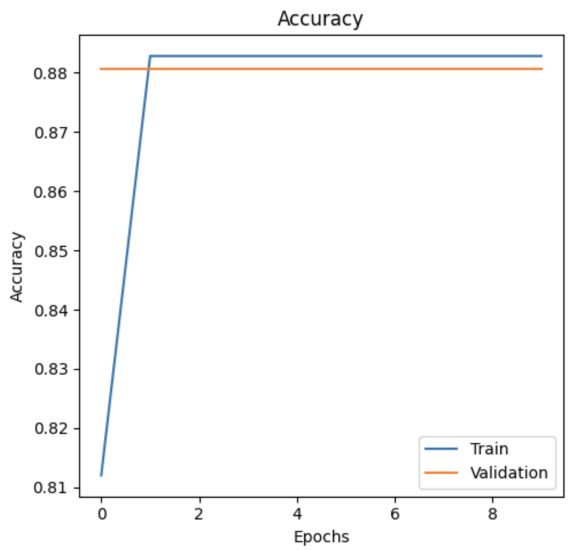


Fig. 3. The accuracy plot over the number of epochs.

Figure 3 illustrates the training and validation accuracies across epochs. The training accuracy initially commences at a baseline of 0, indicating that the model is learning from scratch. As the epochs progress, there's a consistent upward trend in accuracy, reaching a level slightly above 0.88 after just one epoch. This surge hints at the rapid learning capability of the model early on. Subsequently, across the following epochs up to 10, the training accuracy stabilizes, maintaining a consistent performance at this heightened level.

On the other hand, the validation accuracy graphically represents a different scenario. At 0 epochs, the validation accuracy starts impressively high, around 0.88, denoting that the model demonstrates a good understanding of the validation dataset right from the beginning. Remarkably, this level sustains throughout subsequent epochs, remaining stable without significant fluctuation. This stability indicates that the model generalizes well to unseen data, as evidenced by its consistent performance matching the high training accuracy.

This convergence of training and validation accuracies at a relatively high level by the first epoch suggests that the model rapidly grasps the patterns in both datasets, ensuring robust learning and generalization capabilities without overfitting.

Loss

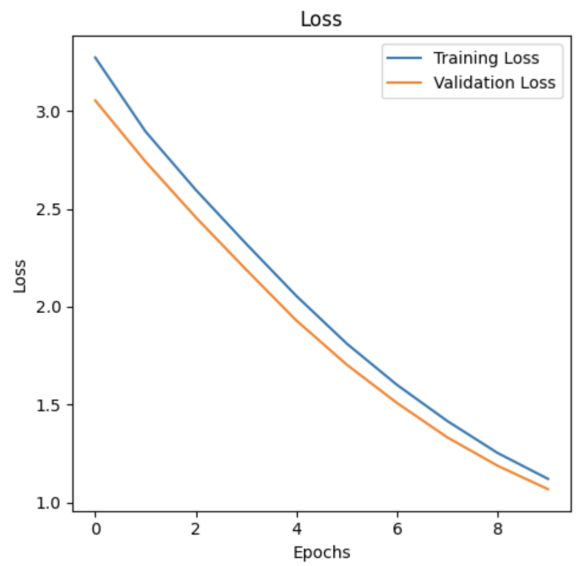


Fig. 4. The loss plot over the number of epochs.

In Figure 4, the trends in training and validation loss are depicted, revealing a pattern of gradual decrease across the epochs. The training loss, initially standing at approximately 3.5, steadily declines with each epoch, indicating that the model is effectively learning from the training data. This decrease continues, reaching a level slightly above 1 after the completion of 10 epochs. This downward trajectory in training loss showcases the model's ability to better fit the training data over time, refining its predictive capabilities.

Simultaneously, the validation loss, which begins around 3, also demonstrates a consistent reduction throughout the epochs. This decrease implies that the model's performance extends beyond the training data and generalizes well to new, unseen data. By the end of the 10 epochs, the validation loss aligns closely with the training loss, hovering around 1. This convergence signifies that the model is learning effectively and doesn't exhibit signs of overfitting, as it performs similarly on both the training and validation datasets.

Both the training and validation loss curves exhibit a smooth, downward trajectory, indicating a progressive improvement in the model's ability to minimize errors. This consistent decline in loss metrics signifies the model's increasing proficiency in capturing patterns and making accurate predictions, reinforcing its reliability and effectiveness in handling both seen and unseen data.

Prediction

In the preparation of input data, the application of tokenizer.texts_to_sequences() method serves to transform textual input into a sequence of integers, effectively encoding the information for compatibility with the model architecture. This step ensures that the model can effectively process and derive meaningful insights from the input text.

In the prediction phase, the utilization of pad_sequences becomes imperative to appropriately pad the input sequence, aligning it with the model's compatibility requirements. This step is crucial for maintaining consistency in the input dimensions, enabling seamless processing within the neural network. Subsequently, the model.predict() function is employed to generate predictions for the output sequence based on the processed input data, leveraging the trained neural network to provide valuable insights and classifications.

Converting the model's output back into human-readable text involves the application of tokenizer.sequences_to_texts(predicted_output_sequence.argmax(axis=-1))[0]. This transformation is pivotal for translating the numerical representation of the predicted output sequence into its corresponding textual form. The decoded output text provides a comprehensible representation of the model's predictions, facilitating the interpretation and evaluation of its performance in real-world applications.

VI. Limitations

Despite our efforts to develop an advanced healthcare chatbot, certain limitations should be noted. The dataset's origin from Hugging Face may introduce biases or constraints from the collected conversations, potentially impacting the model's performance, especially in scenarios not well-covered. Additionally, while the model's performance metrics are promising, they might not entirely reflect its real-world efficacy, influenced by factors like network latency and variations in user input within live healthcare environments.

VII. Future Study

To address these limitations and further improve the chatbot's capabilities, future studies should focus on augmenting the dataset with a wider range of medical scenarios and patient queries, enhancing the model's adaptability and generalizability. Incorporating advanced NLP techniques beyond word embedding, such as transformer models like BERT or GPT, could significantly enhance the chatbot's contextual understanding. Additionally, integrating real-time learning mechanisms and conducting comprehensive user studies within live healthcare settings would provide invaluable insights for refining the chatbot's performance and user experience.

VIII. Conclusion

This advanced chatbot represents a significant stride in enhancing communication and elevating the standard of patient care. At its core, the model is engineered to provide accurate and personalized medical assistance, marking a pivotal shift in the landscape of healthcare interactions.

The success of this chatbot is underpinned by a meticulously crafted model that seamlessly navigates through key stages of development. Through meticulous model building, the architecture is tailored to intelligently process and comprehend input text, laying the groundwork for an unparalleled level of precision in medical guidance. This precision is further bolstered by a robust compilation process, leveraging state-of-the-art optimization algorithms and loss functions to refine the model's capacity for accurate predictions.

In action, the chatbot's predictive prowess comes to the forefront. Through a series of predictive steps, it transforms input text into tailored medical recommendations, aligning its responses with the specific needs and inquiries of the user. This transformative capability not only streamlines communication but also delivers a level of personalized care that transcends traditional healthcare interfaces.

In summary, the chatbot, with its adept model architecture, demonstrates a paradigm shift in healthcare communication. By seamlessly integrating model building, compilation, and prediction steps, it stands as a beacon of efficiency, converting raw input text into precise and invaluable medical guidance. This innovation represents a cornerstone in the ongoing evolution of patient care, showcasing the potential for artificial intelligence to revolutionize and elevate healthcare interactions to unprecedented heights.

IX. Sources

Our Dataset Link

https://huggingface.co/datasets/lavita/ChatDoctor-HealthCareMagic-100k

Github Link for our Code

https://github.com/Kunal-1669/deep_learning_Final_Project.

References

[1] Decker, H., Trang, K., Ramirez, J., Colley, A., Pierce, L., Coleman, M., Bongiovanni, T., Melton, G. B., & Wick, E. (2023). "Large Language Model-Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures." JAMA Network Open, 6(10), e2336997. https://doi.org/10.1001/jamanetworkopen.2023.36997

[2] Erdogan, C., Buluş, H. N., & Diri, B. (2013). "Analyzing the performance differences between pattern matching and compressed pattern matching on texts." In 2013 International Conference on Electronics, Computer and Computation (ICECCO) (pp. 135-138). Ankara, Turkey. https://doi.org/10.1109/ICECCO.2013.6718247

[3] Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013). "Using deep learning to enhance cancer diagnosis and classification." In Proceedings of the International Conference on Machine Learning.

[4] Hwang, T.-H., Lee, J., Hyun, S.-M., & Lee, K. (2020). "Implementation of interactive healthcare advisor model using chatbot and visualization." In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 452-455). Jeju, Korea (South). https://doi.org/10.1109/ICTC49870.2020.9289621.

[5] Kandpal, P., Jasnani, K., Raut, R., & Bhorge, S. (2020). "Contextual Chatbot for Healthcare Purposes (using Deep Learning)." In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) (pp. 625-634). London, UK. https://doi.org/10.1109/WorldS450073.2020.9210351.

[6] Maher, S. K., Bhable, S. G., Lahase, A. R., & Nimbhore, S. S. (2022). "AI and Deep Learning-driven Chatbots: A Comprehensive Analysis and Application Trends." In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 994-998). Madurai, India. https://doi.org/10.1109/ICICCS53718.2022.9788276.

[7] Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study." Digital Health, 5. https://doi.org/10.1177/2055207619871808.

[8] Rajesh, V., Perumal, B., Ganesh, U. S., Rajkumar, V., Kumar, D. M., & Kumar, M. (2023). "Building Customer Support Chatbots With Intent Recognition." In 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN) (pp. 1-5). Vellore, India. https://doi.org/10.1109/ViTECoN58111.2023.10157329.

[9] Rodríguez, F. M., Vilares, J., Alonso Pardo, M., & Vilares Ferro, M. (2002). "Tokenization and Proper Noun Recognition for Information Retrieval." Database and Expert Systems Applications, International Workshop on, 246. https://doi.org/10.1109/DEXA.2002.1045906.

[10] Siddique, Sarkar, & Chow, J. C. L. (2021). "Machine Learning in Healthcare Communication." Encyclopedia, 1(1), 220-239. https://doi.org/10.3390/encyclopedia1010021.

[11] Ambekar, S., & Phalnikar, R. (2018). "Disease Risk Prediction by Using Convolutional Neural Network." In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). Pune, India. https://doi.org/10.1109/ICCUBEA.2018.8697423.

[12]Xu, L., Sanders, L., Li, K., & Chow, J. (2021). "Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review." JMIR Cancer, 7(4), e27850. https://doi.org/10.2196/27850.