Grenze International Journal of Engineering and Technology, Jan Issue

A Survey of Machine Learning Models used in Development of Medical Chatbots

Priyank Shah[1], Manik Pratapwar[2], Suraj Padia[3] and Dr. Sanjay T. Gandhe[4]
[1-4]SCTR's Pune Institute of Computer Technology/Electronics and Telecommunication, Pune, India
Email: piyushah2903@gmail.com, {pratapwar999, surajpadia9404}@gmail.com,principle@pict.edu

***Abstract*— A new era of patient care and health care has started because of the integration of artificial intelligence (AI) and machine learning (ML) technologies. This paper analyzes significant developments achieved in the fieldof medical chatbots, which are now efficient instruments for improving patient engagement, increasing access to healthcare, and accelerating medical operations. The capacity of medical chatbots, powered by AI and ML algorithms, to offer timely and individualized healthcare information, symptom assessment,and even early diagnoses, has been established. They use naturallanguage processing (NLP) methods to converse in human-like ways, making them accessible for individuals with different degrees of digital literacy. Additionally, they can be integrated with other healthcare systems and electronic health records (EHRs), facilitating seamless exchange of data while improving the client's experience.**

***Index Terms*— Sequence to Sequence Models (Seq2Seq), Naive Bayes (NB), Long Short Term Memory (LSTM), Recurrence Neural Network(RNN)**

I. INTRODUCTION

In the realm of modern technology, Chatbots have emergedas dynamic systems that seamlessly engage with human users through natural language interactions. The vast reservoir ofonline data empowers Chatbots to deliver precise and efficient information tailored to the unique needs of each user. In this context, we introduce our envisioned Medical Chatbot, designed to interact with users and provide them with an in- telligent, akin-to-professional medical consultation experience. This Chatbot adeptly analyzes initial messages, extracting keywords that hint at potential medical concerns expressed by the user.

While existing Medical Chatbots do exist, their primary function is not to prescribe medications for specific ailments but rather to connect users with a Medical Forum and display similar queries related to their reported symptoms, allowing for insights that doctors may have previously provided. It is worth noting that several limitations plague current systems [12]. Instant responses to patients are often lacking, as individuals must endure extended waiting times for specialist feedback. Additionally, certain processes may incur costs for live chator telecommunication with online doctors.

Upon initiating the application, the user is greeted by theChatbot. After completing a preliminary form, the user's inputis processed by the Chatbot, which subsequently validates thesubmitted details. With user agreement, the medical inquiryproceeds. The Chatbot engages the user with further questionsto gain a comprehensive understanding of the patient's query.The user responds to the Chatbot's inquiries accordingly, and the Chatbot generates a well-informed response. If the userfinds the provided information satisfactory, they can proceed

Grenze ID: 01.GIJET.10.1.37_1

with confidence in their interaction with the Chatbot. Thisfusion of technology and healthcare offers users a valuable toolfor gaining medical insights and making informed decisions about their well-being.

II. APPLICATION OF MACHINE LEARNING MODELS IN DEVELOPMENT OF CHATBOTS

Chatbots are a form of human-computer dialog system that operates through natural language processing using text or speech, chatbots are automated and generally run 24/7. Itis mainly used to drive conversion and designed to handle millions of requests at a time.

A group of intelligent, conversational software algorithms called chatbots are triggered by input in natural language. They have the capacity to understand commands, comprehend input, and carry out tasks. Even though chatbots have been around for a while, they are becoming more advanced becauseof the availability of data, increased processing power, and open-source development frameworks. These elements have prompted the widespread use of chatbots across a varietyof sectors and domains. We frequently come with chatbotsin a variety of settings, from customer service, social mediaforums, merchant websites to availing banking services, alike.In general, chatbots are made to accomplish specific tasks. For instance, customer care chatbots are created specifically tomeet the needs of customers who request assistance, whereasconversational chatbots are created to engage in conversationwith users. It is really possible to train with a large datasetand archive human level interaction, but organizations have torigorously test and check their chatbot before releasing into production [14].

A. Machine Learning Models

Several algorithms are used in developing AI chatbots. Among all of them, natural language processing-based algorithms are widely used. As chatbot gets input in natural language, so text processing, classification and interpretation becomes important when it comes to quality chatbots.

Popular chatbot algorithms include the following ones are Naïve Bayes Algorithm,Vector Machine Natural language processing (NLP)Recurrent neural networks (RNN) Long short-term memory (LSTM).

III. MACHINE LEARNING MODELS IN DEPTH

A. Naive Bayes

Naive Bayes is a machine learning algorithm that is commonly used in natural language processing (NLP) tasks, including chatbot development. It is based on the idea of using Bayes' theorem to make predictions about the likelihood of certain events occurring, based on the presence or absence of certain features.

In the context of chatbots, Naive Bayes can be used to classify user messages into predefined categories or intents. For example, a chatbot might use Naive Bayes to classify user messages as being related to weather, sports, news, or some other topic. To do this, the chatbot would need to be trainedon a dataset of user messages, along with labels indicatingthe intent of each message. The chatbot would then use this training data to learn the features that are associated with each intent and use these features to make predictions about the intent of new, unseen messages.

The paper [25] utilizes the Naive Bayes method to enablea conversational interface for medical inquiries through a web application, accommodating both voice and text interactions. This system encourages open-ended conversations between doctors and patients, allowing for natural expressions of health-related queries without the need for physical clinic visits. The Naive Bayes algorithm processes these queries, delivering human-like responses that are tailored to the questions asked. Powered by natural language processing technology, the bot can comprehend sentences with grammatical errors or lacking clarity, enhancing its user-friendliness. It employs acombination of retrieval and generation methods to provideanswers, with a focus on presenting information in a clear and analytical manner. Ultimately, this project aims to make healthcare information readily accessible and understandable, fostering a seamless exchange of medical knowledge between users and the system.

According to [25] Bayesian network is equivalent to a naive Bayes classifier. A crucial problem in data mining and machine learning is classification. A learning algorithm's objective in classification is to build a classifier from a set of training examples with a class label. The simplest type of Bayesian networkis the naive Bayes, in which all attributes are independent ofone another regardless of the value of the class variable. We refer to this as conditional independence. In most real-world applications, it is evident that the conditional independence assumption is erroneous only sometimes. Increasing the structure of naive Bayes to explicitly reflect attribute relationships is an easy way to overcome this constraint. [26]In study of[27], three separate optimization models which take class probabilities and conditional probabilities into accountas unknown variables for the naive

Bayes classifier. The best values for these variables were then determined using optimization techniques. On 14 real-world binary classification data sets, comparison of suggested models to NB, TAN, the SVM, C4.5, and 1-NN. using a median-based discretization approach and the SOAC algorithm when combined with the Fayyad and Irani method, the values of features in data sets are discretized. Results of numerical experiments were given. The findings show that the suggested models outperform NB, TAN, the SVM, C4.5, and 1-NN in terms of accuracy while still retaining NB's straightforward structure.

B. Sequence to Sequence Models

Seq2seq was first introduced for machine translation, by Google. Before that, the translation worked in a very naïve way. Each word that you used to type was converted to itstarget language giving no regard to its grammar and sentence structure. Seq2seq revolutionized the process of translation by making use of deep learning. It not only takes the current word/input into account while translating but also its neigh- borhood. Seq2Seq (Sequence-to-Sequence) is a type of model in machine learning that is used for tasks such as machine translation, text summarization, and image captioning. The model consists of two main components:
Encoder Decoder Seq2Seq models are trained using a dataset of input-output pairs, where the input is a sequenceof tokens, and the output is also a sequence of tokens. Themodel is trained to maximize the likelihood of the correct output sequence given the input sequence.
It is a component of machine learning (ML) and is mostly used in models for language processing. Recurrent neural net- works (RNNs) can be used to implement it utilizing encoder- decoder based machine interpretation, which converts input sequences into successive output sequences with tags and consideration values. The goal is to use two RNN that willwork together with a special token to try and predict the next state arrangement based on the previous succession. Attention- based models and connectionist temporal classification (CTC) are other ways to create sequence-to-sequence models. To train the model with just one command, Google introduced the sequence-to-sequence model for machine translation [1]. Model's name was also created to be easily expandable and repeatable, the code files were arranged methodically and made it simple to develop upon and duplicate.
Paper [29] implemented a careful research strategy in the assessment of three sequence to sequence models: recurrence neural network (RNN), attention-based model, connectionist temporal classifications (CTC).

1) Connectionist temporal classification

The CTC algorithm that [30] suggested. With the help of this approach, start- to-finish models can be created without the need for a case- level organization of the objective names in the preparatory articulation. Recurrent neural networks, sometimes referred to as encoders, are predicted to be used in CTC to specify the likelihood of the output condition [31]. Since" CTC introduces a special blank label and allows for repetition of labels to force the output and input sequences to have the same length," it also uses the sequence-to-sequence method to help address any issues related to the length of the output labels when it is shorter than the length of the input sequence. Blank symbols frequently predominate in CTC outputs [32]. This is a huge benefit for CTC. When applying sequence-to-sequence models to various tasks, including translation.

2) Recurrence Neural Network

According to [29] Sequence to sequence modeling has been synonymous with recurrent neural network-based encoder-decoder architectures [28]. The encoder RNN processes an input sequence x of m elements and returns state representations. The decoder RNN takes z and generates the output sequence left to right,one element at a time. To generate output, the decoder computes a new hidden state based on the previous state, an embedding giof the previous target language word, aswell as a conditional input derived from the encoder output. Based on this generic formulation, various encoder-decoder architectures have been proposed, which differ mainly in the conditional input and the type of RNN.

3) Attention model

Like the RNN transducer model, it is a consideration-based model that includes an encoder organize. However, a consideration-based model uses a single decoderto deliver an appropriation over the marks modelled on the full grouping of historical forecasts and the acoustics, in contrast to the RNN transducer, in which the encoder and the expectation arrange are displayed autonomously and combined in the joint system [31].
After assessment of paper [29],[30],[31] notably, the workof [31] compared the three different approaches and found the most promising approach to be" the RNN transducer, attention-based models, and a novel RNN transducer augmented with attention."

IV. RECURRENCE NEURAL NETWORK AND LONG SHORT-TERM MEMORY

A. Recurrence Neural Network

The concept of sequence-to-sequence models employing the RNN approach makes use of two RNN that will collaboratewith a special token and try to predict the next state arrangement from the previous succession. In contrast to the CTCalignment model, the RNN transducer uses a distinct repeat lease expectation arrangement over the output sequences. The encoder can be conceptualized instinctively as an auditory model, whereas the expectation arrangements are essentiallycomparable to a language model. The expectation arrange receives an input vector and processes it in accordance withthe entire label sequence [31].

Humans don't start their thinking from scratch every second. As you read this essay, you understand each word based on your understanding of previous words. You don't throweverything away and start thinking from scratch again. Your thoughts have persistence.

Traditional neural networks can't do this, and it seems like a major shortcoming. For example, imagine you want to classify what kind of event is happening at every point in a movie.It's unclear how a traditional neural network could use its reasoning about previous events in the film to inform laterones.

Recurrent neural networks address this issue. They arenetworks with loops in them, allowing information to persist. In the above diagram, a chunk of neural network, A, looksat some input xt and outputs a value ht. A loop allows information to be passed from one step of the network to the next.

These loops make recurrent neural networks seem kind of mysterious. However, if you think a bit more, it turns out that they aren't all that different than a normal neural network. A recurrent neural network can be thought of as multiple copiesof the same network, each passing a message to a successor. Consider what happens if we unroll the loop:

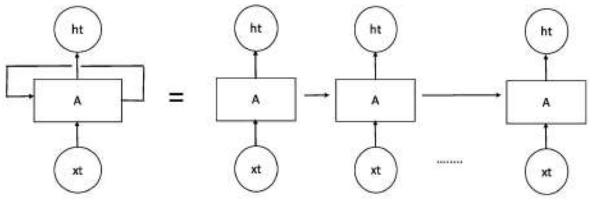

Fig. 2. An unrolled recurrent neural network

In Fig.2 chain-like nature reveals that recurrent neural networks are intimately related to sequences and lists. They're the natural architecture of neural networks to use for such data.

B. Long Short-Term Memory

Long Short-Term Memory networks, commonly referred to as LSTMs, represent a remarkable advancement in the field of artificial neural networks. These powerful models have demonstrated exceptional performance on challenging machine learning tasks, particularly when dealing with ex- tensive labeled training datasets. However, there has beena longstanding limitation to their application: LSTMs are inherently designed for tasks where the input and output arenot sequences of data.

According to [34] The goal of the LSTM is to estimate the conditional probability for input sequence for its corresponding output sequence whose length may differ. The Problem of Long-Term Dependencies one of the appeals of RNNs is the idea that they might be able to connect previous information to the present task, such as using previous video frames might inform the understanding of the present frame. If RNNs could do this, they'd be extremely useful. But can they? It depends.

Sometimes, we only need to look at recent information to perform the present task. For example, consider a languagemodel trying to predict the next word based on the previous ones. If we are trying to predict the last word in "the cloudsare in the sky," we don't need any further context – it's pretty obvious the next word is going to be sky. In such cases, where the gap between the relevant information and the place that it's needed is small, RNNs can learn to use the past information.

But there are also cases where we need more context. Consider trying to predict the last word in the text "I grewup in France. . . I speak fluent French." Recent informationsuggests that the next word is probably the name of a language, but if we want to narrow down which language, we need the context of France, from further back. It's entirely possible for the gap between the relevant information and the point whereit is needed to become very large.

Unfortunately, as that gap grows, RNNs become unable to learn to connect the information.

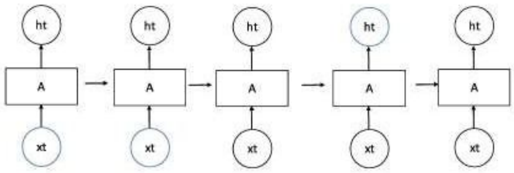

Fig. 3. An unrolled recurrent neural network

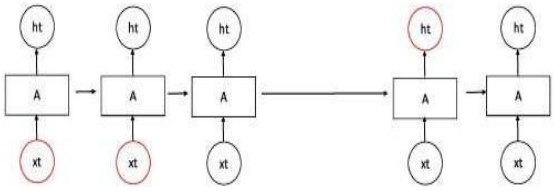

Fig. 4. An unrolled recurrent neural network.

LSTMs are explicitly designed to avoid the long-term de- pendency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn.

All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as asingle tanh layer.

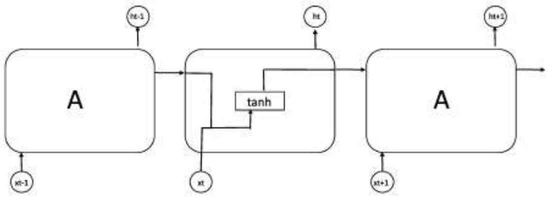

Fig. 5. The repeating module in a standard RNN contains a single layer.

LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a singleneural network layer, there are four interacting in a very special way.

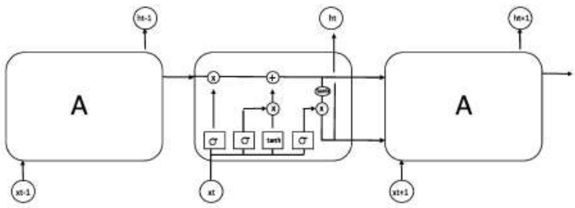

Fig. 6. The repeating module in an LSTM contains four interacting layers.

V. Conclusion

In conclusion, Long Short-Term Memory (LSTM) networks have demonstrated their superiority over traditional Recur- rent Neural Networks (RNNs) and Naive Bayes classifiers in various machine learning and natural language processing tasks. LSTMs excel in capturing long-range dependencies, making them particularly well-suited for sequential data and time series analysis. Their ability to maintain and update amemory cell allows them to handle sequences of varying lengths and effectively mitigate the vanishing gradient problem that plagues standard RNNs.

Furthermore, LSTMs have been instrumental in the successof many state-of-the-art models in areas such as speech recognition, machine translation, sentiment analysis, and more. Their capacity to learn complex patterns and contextual information over extended sequences has made them the to choose for tasks requiring an understanding of temporal dynamics.

Comparatively, Naive Bayes classifiers, while simple andinterpretable, often assume feature independence, which can limit their performance in scenarios with correlated features or intricate dependencies. While Naive Bayes can serve as a baseline model for some classification problems, it generally lacks the capacity to model complex relationships present in sequential or high-dimensional data.

In summary, LSTM networks have proven to be a morecapable and versatile choice when compared to RNNs and Naive Bayes classifiers, especially for tasks involving sequential data and complex dependencies. However, the choice of model ultimately depends on the specific problem at hand and the trade-offs between complexity, interpretability, and performance.

TABLE I. COMPREHENSIVE REVIEW OF RESEARCH PAPERS

Project	Author	Techniques	Results	Conclusions
An intelligent Chatbot using deep learning with Bidirectional RNN and attention model [2]	Manyu Dhyani⁎ and Rajiv Kumar	Neural Machine Translation (NMT) model ,BRNN (BRNN),NMT model	Bleu score-30.16	Learning rate is useful but only when model is trained with large data and for longer period of time. For limited period of time or with less data, there isno significant change.
Diabot: A Predictive Medical Chatbot using Ensemble Learning[3]	Manish Bali, Samahit Mohanty, Subarna Chatterjee, Manash Sarma, Rajesh Puravankara	Ensemble classifier	Generic Diseases-86, Diabates-84.2	The larger and complex the dataset, the better the accuracy to implement it in a real-time scenario.
Dr.Vdoc: A medical chatbot that acts as a virtual doctor[4]	SK Mishra, D Bharti, N Mishra	NLP and pattern matching algorithm.	80% of accuracy.	The full consultation is given at a free of cost, which is not a good practice.
Medchatbot:An UMLS based chatbot for medical student[5]	Hameedullah kazi,B.S. Chowdhry,Zeesha Memon	AIML pattern technique for Pattern matching.	47% of accuracy.	The system has not been specially designed for the task of supporting natural dialog in chatbots.
Towards a chatbot for digital counselling [6]	Gillian Cameron ,David Cameron,Gavin Megaw,Raymond Bond,Maurice Mulvenna,SiobhanO'Neill	NLP and pattern matching algorithm.	60% of accuracy	The issue here is to maintain the ethical considerations.
Pharmabot: A pediatric generic Medicine consultant Chatbot[7]	Benilda Eleonor V. Comendador, Michael B. Francisco, Jefferson S. Medenilla, Sharleen Mae T.Nacion, and Timothy Bryle E. Serac	Left and Right Parsing Algorithm.	2.1923, is less than the table value of 2.447 with 0.05 level of significance for two-tailed test and 6 as degrees of freedom	It can be developed into a web-based application so that everyone can access and use it.
Chatterbot implementation[8]	Sandeep A. Thorat, Vishakha Jadhav	Transfer Learning and LSTM Encoder-Decoder Architecture	RMSE for training dataset = 6.42 RMSE for testing dataset =18.24	They are more complicated than traditional RNNs and require more training data in order to learn effectively.
They are more complicated than traditional RNNs and require more training data in order to learn effectively.	Pavlidou Meropi; Antonis S. Billis; Nicolas D. Hasanagas; Charalambos Bratsas; Ioannis Antoniou; Panagiotis D. Bamidis	Natural Language Processing.	Conditional Entropy 96.77419% Deep NN-82% RBF Kernel SVM-54.01%	The issue here is,it requires a compact Adjacency Matrix based on the dialogues.
Medical Chatbot using Support Vector Machine Learning Algorithm[11]	B. Tamizharasi1, L.M. Jenila Livingston1 and S. Rajkumar1	Support Vector Machine	SVM Classifier-92% KNN- 87% NaiveBayes-81%	By taking the advantages of the SVM algorithm medical chatbots can be extended and useddeeply with other medical systems where predictions can be done. It can be further extended toschedule doctor visits and remind patients of a next appointment or routine check-up.

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.

[2] Dhyani, M. and Kumar, R., 2021. An intelligent Chatbot using deep learning with Bidirectional RNN and attention model. Materials today: proceedings, 34, pp.817-824.

[3] Bali, M., Mohanty, S., Chatterjee, S., Sarma, M. and Puravankara, R., 2019. Diabot: a predictive medical chatbot using ensemble learing. International Journal of Recent Technology and Engineering, 8(2), pp.6334-6340.

[4] Mishra, S.K., Bharti, D. and Mishra, N., 2017. Dr. Vdoc: a medical chatbot that acts as a virtual doctor. Journal of Medical Science and Technology, 6(3).

[5] Kazi, H., Chowdhry, B.S. and Memon, Z., 2012. MedChatBot: an UMLS based chatbot for medical students.

[6] Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C. and McTear, M., 2017, July. Towards a chatbot for digital counselling. In Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017). BCS Learning Development.

[7] Comendador, B.E.V., Francisco, B.M.B., Medenilla, J.S. and Mae, S., 2015. Pharmabot: a pediatric generic medicine consultant chatbot. Jour- nal of Automation and Control Engineering, 3(2).

[8] Thorat, S.A. and Jadhav, V., 2020, April. A review on implementation issues of rule-based chatbot systems. In Proceedings of the international conference on innovative computing communications (ICICC).

[9] Borah, B., Pathak, D., Sarmah, P., Som, B. and Nandi, S., 2019. Survey of textbased chatbot in perspective of recent technologies. In Computational Intelligence, Communications, and Business Analytics: Second International Conference, CICBA 2018, Kalyani, India, July 27–28, 2018, Revised Selected Papers, Part II 2 (pp. 84-96). Springer Singapore.

[10] Meropi, P., Billis, A.S., Hasanagas, N.D., Bratsas, C., Antoniou, I. and Bamidis, P.D., 2017, June. Conditional entropy-based retrieval model in patient-carer conversational cases. In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 576- 581). IEEE.

[11] Tamizharasi, B., Livingston, L.J. and Rajkumar, S., 2020, December. Building a medical chatbot using support vector machine learning algorithm. In Journal of Physics: Conference Series (Vol. 1716, No. 1, p. 012059). IOP Publishing.

[12] Lokman, A.S., Zain, J.M., Komputer, F.S. and Perisian, K., 2009, October. Designing a Chatbot for diabetic patients. In International Con- ference on Software Engineering Computer Systems (ICSECS'09) (pp. 19-21).

[13] Rarhi, K., Bhattacharya, A., Mishra, A. and Mandal, K., 2017. Auto- mated medical chatbot. Available at SSRN 3090881.

[14] Blanc, C., Bailly, A., Francis, E´ ., Guillotin, T., Jamal, F., Wakim, B. and Roy, P., 2022. FlauBERT vs. CamemBERT: Understanding patient's answers by a French medical chatbot. Artificial Intelligence in Medicine, 127, p.102264.

[15] Safi, Z., Abd-Alrazaq, A., Khalifa, M. and Househ, M., 2020. Tech- nical aspects of developing chatbots for medical applications: scoping review. Journal of medical Internet research, 22(12), p.e19127.

[16] Mehfooz, F., Jha, S., Singh, S., Saini, S. and Sharma, N., 2021. Medical chatbot for novel COVID-19. In ICT Analysis and Applications: Proceedings of ICT4SD 2020, Volume 2 (pp. 423-430). Springer Singapore.

[17] Bao, Q., Ni, L. and Liu, J., 2020, February. HHH: an online medical chatbot system based on knowledge graph and hierarchical bi-directional attention. In Proceedings of the Australasian computer science week multiconference (pp. 1-10).

[18] Athota, L., Shukla, V.K., Pandey, N. and Rana, A., 2020, June. Chatbot for healthcare system using artificial intelligence. In 2020 8th International conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO) (pp. 619-622). IEEE.

[19] Bulla, C., Parushetti, C., Teli, A., Aski, S. and Koppad, S., 2020. A review of AI based medical assistant chatbot. Research and Applications of Web Development and Design, 3(2), pp.1-14.

[20] Kalla, D. and Samiuddin, V., 2020. Chatbot for medical treatment using NLTK Lib. IOSR J. Comput. Eng, 22.

[21] Mittal, M., Battineni, G., Singh, D., Nagarwal, T. and Yadav, P., 2021. Web-based chatbot for frequently asked queries (FAQ) in hospitals. Journal of Taibah University Medical Sciences, 16(5), pp.740-746.

[22] Liu, Z., Li, X., You, Z., Yang, T., Fan, W. and Yu, P., 2021, July. Medical triage chatbot diagnosis improvement via multi-relational hyperbolic graph neural network. In Proceedings of the 44th International ACM SIGIR conference on research and development in information retrieval (pp. 1965-1969).

[23] Hossain, M.M., Pillai, S.K., Dansy, S.E., Bilong, A.A. and Panessai, I.Y., 2021. Mr. Dr. Health-Assistant Chatbot. International Journal of Artificial Intelligence, 8(2), pp.58-73.

[24] Angappan, A.K., Khan, N.J., Sharan, M. and Ahamed, S.F., 2021. AI- based healthcare chatbot system.

[25] Behera, B., Chappie—A semi-automatic Intelligent Chatbot(pp. 1–5).

[26] Revathy, S., 2020, June. Health care counselling via voice Bot using multinomial naive bayes algorithm. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1063-1067). IEEE.

[27] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[28] Taheri, S. and Mammadov, M., 2013. Learning the naive Bayes classifier with optimization models. International Journal of Applied Mathematics and Computer Science, 23(4), pp.787-795.

[29] Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. Advances in neural information process- ing systems, 27.

[30] Gehring, J., Auli, M., Grangier, D., Yarats, D. and Dauphin, Y.N., 2017, July. Convolutional sequence to sequence learning. In International conference on machine learning (pp. 1243-1252). PMLR.
[31] Battenberg, E., Chen, J., Child, R., Coates, A., Li, Y.G.Y., Liu, H., Satheesh, S., Sriram, A. and Zhu, Z., 2017, December. Exploring neural transducers for end-to-end speech recognition. In 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 206- 213). IEEE.
[32] Huang, X., Qiao, L., Yu, W., Li, J. and Ma, Y., 2020. End-to-end sequence labelling via convolutional recurrent neural network with a connectionist temporal classification layer. International Journal of Com- putational Intelligence Systems, 13(1), pp.341-351.
[33] Huang, Z., Xu, W. and Yu, K., 2015. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991.
[34] Gers, F.A., Schmidhuber, J. and Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), pp.2451- 2471.