MedBot: A Medical Chatbot for Disease Detection and Suggestion Through Machine Learning Models

Avina Wakchaure[1], Shamli Kavle[2], Shruti Surve[3], Apurva Patil[4], Meenal Nerkar[5]
[1,2,3,4]*Department of Computer Engineering, AISSMS Institute of Information Technology, Pune*
[5]*Project Guide, Department of Computer Engineering, AISSMS Institute of Information Technology*

***Abstract* — The Healthcare sector is one of the fastest growing and highly competitive sector that has been improving with the insertion of technology into this paradigm. The healthcare sector has been facing a severe lack of medical professionals that has been evident in the recent pandemic. The recent pandemic exposed the dearth of doctors and other staff in the hospitals and clinics. This is a problematic issue that will require a lot of effective improvements, but the lack of the professionals has been an issue that has lasting consequences right away. The doctors that are consulting are overburdened with the patients and the elderly individuals have a hard time travelling to the hospital every time for a consultation. Therefore, the paradigm of remote diagnosis comes to the rescue. For this purpose, an effective approach called MedBot is proposed in this research article that utilizes K Nearest Neighbor clustering and Linear Regression along with Artificial Neural Networks and Decision Making. The approach has been effectively evaluated for its disease prediction and has resulted in extremely satisfactory results.**

Keywords: K nearest Neighbors, Linear Regression, Artificial Neural Networks and Decision Tree

I INTRODUCTION

Healthcare sector is one of the most essential as well as the most prioritized sector in the entire world. This is due to the nature of the sector that allows for the optimal functioning of the human being. The health is one of the central aspects of an individual’s well-being. It is also one of the most researched aspect that goes through immense technological advancements ever year. This is highly useful for improving the overall living standards of humans. The healthcare sector is tasked with the recovery and handling of individuals that are afflicted with disease and other infections or maladies. The medical community has always been tirelessly working towards the eradication of the pain and suffering of the patients.

There has always been a lack of medical professionals and other healthcare professionals across the globe. This has been evident in the recent pandemic that had led to an immense shortage of staff in the hospitals and clinics. This exposed the fact that there a lot of medical centers are understaffed which is a very problematic occurrence. This can be actually understood as the process to become a doctor is highly complex and a very difficult task that is highly competitive. This leads to a lot less individuals opting to take up medicine and even fewer that get selected for it. Therefore, the medical sector is always understaffed leading to increased workload to almost all the doctors all over the world.

The lack of the medical professionals has been one of the most problematic occurrences that has been one of the biggest concerns since it was highlighted by the recent pandemic. The process of improving the numbers of medical professionals is a long and arduous process which has been undertaken by nations, India itself has been increasing the medical institutions across the country to allow for more students to take up the medical field. But this will take time and there is constant influx of patients and other individuals that require the assistance of a medical professional right now. This increasing number of patients has been putting a lot of pressure on the existing professionals that are overworked and exhausted.

Some of the patients are also suffering due to the fact that there are a number of different patients with problems that does not allow them to travel large distances to the hospital or the clinic. This is also true to elderly individuals that cannot reach the hospital due to their old age which can be quite problematic. Therefore, there has been an increased interest of individuals towards the development of remote diagnosis approaches that can alleviate the problems of the professionals as well as the patient. The remote diagnosis has been one of the hottest topic in the recent researches which has been getting increased attention in the recent years. Therefore, this research article defines an effective approach for Medical chatbot through the use of Machine learning methodologies.

In recent years, tremendous progress has been experienced in research on Artificial Intelligence that partially propagated the growth of sophisticated chatbots. Their capacity to simulate human-like discussions with a user via voice, text, smart display as well as multisensory communication makes them more popular. Unlike static apps, user intentions and choices may be understood and communicated through interfaces. Such technology quickly gains popularity in the field of medicine, where there is insufficient expert care. Chatbots are capable of providing a more cost-effective and accessible alternative.

The Machine Learning technology has thus been chosen for this purpose as it offers great precision. One of the most efficient machine learning techniques and is a member of the large and competent family of neural networks that is the artificial neural network. This technique can make a medical chatbot incredibly effective and very precise in its diagnosis. Therefore, an effective and useful application of a medical chatbot has been envisioned in this survey article, the related researches have been crucial in the development of an effective strategy that will be discussed in the upcoming research articles.

This research study has five sections: section 2 analyzes previous work as a literature survey, section 3 thoroughly explains the suggested approach, section 4 assesses the system's performance, and section 5 finishes the paper with suggestions for future improvements.

II RELATED WORKS

X. Ren et al. [1] presented an experimental study using conversational agent-based interactions to improve intelligent decisions and help during healthcare consultation. ConsultAI, an interactive chatbot assistant meant to aid the occupational health physician in real-time, was utilized to carry out the recommended conversational strategy. The authors performed field research with eight occupational health consultations to establish the following: the usefulness of ConsultAI in the context of occupational health, and the influence of ConsultAI chatbot interaction styles on the user experience. According to the quantitative findings, physicians evaluated ConsultAI's conversational interface favorably in terms of information dependability and technology adoption. The authors also observed that the on-demand interaction style of the chatbot was favored over the proactive engagement strategy.

B. Zhang et al. proposed a CART framework for BP prediction depending on biological feature data from the CM400 health monitor, such as ECG, PTT, PPG, SPO2, and HR. To avoid model overfitting, the model's optimum parameters were found using the cross-validation technique. In a matrix of accuracy rate, deviation rate, RMSE, and TIC, the CART framework was compared to several traditional methodologies like linear regression, SVM, ridge regression, and neural networks to validate its effect. To discover which variables were the most associated, the Pearson correlation coefficient was used [2]. The CART framework beat the other four models in the testing, with a prediction rate of more than 90%.

G. Mao and colleagues propose a hierarchical multi-presentation aggregation network for multi-turn answer selection in retrieval-based chatbots. For hierarchical aggregation, the authors construct the self-aggregation and matching aggregation mechanisms. Two approaches merge multi-grained representations step by stage, reducing redundancies and distilling high-level information. The authors regard the candidate's answer to be a real element of the context and improve the model framework by including it [3]. Their approach beats state-of-the-art models in two large-scale response selection data sets, according to experimental findings. To demonstrate that their model can capture essential information for response selection, they offer a visualization result. Then, to study the influence of each module, they undertake ablation analyses, and the resulting findings confirm their utility and efficiency. The authors also look at how the model's performance is affected by the number of turns in the context.

Health Assistant Bot, a Telegram-based conversational assistant for assisting patients in their everyday tasks, was introduced by M. Polignano et al. The agent was built with a modular design in mind so that additional functionality may be readily added as needed. Users may track their therapies, and biological parameters, receive doctor recommendations and self-diagnose using the system. The dialogue is carried out via a text-based interface, which makes the interaction easy while also being error-resistant [4]. The interface, gateway, and server-side operations are the three primary aspects of the proposed platform's design. Each of them is self-contained to ensure strong internal coherence and little overlap with the other modules' functions.

In response to the demands of data consumers who are responsible for exploring ways to respond to the ongoing pandemic, R. Oruche et al. described a unique evidence-based recommender system called KnowCOVID-19. They specifically address the topic of workflow constraints they encounter while conducting literature reviews involving the constantly developing COVID-19 published databases. Their unique KnowCOVID-19 system enables data consumers to leverage edge computing services with recommender modules for publications analytics based on the Levels of Evidence Pyramid using thin clients [5]. The authors suggested category framework, which employs the LDA inference algorithm to filter articles depending on their clinical category in the evidence-based practice standards, is also utilized in the evidence-based filtering strategy. Furthermore, the KnowCOVID-19 offers a social plane-based follow-on social filtering that lets data users publish/subscribe to significant insights derived from the automated literature review procedure.

To identify mortality during follow-up, G. N. Ahmad et al. devised a prediction technique depending on hybrid intelligent machine learning. The technique was evaluated using data from a database of heart disease clinical records. One of the most difficult challenges in medicine is forecasting illness sickness by selecting significant factors. To predict cardiac sickness, researchers utilized a variety of algorithms, as well as the feature selection technique SFS. For verification, the framework employs a K-fold cross-validation technique [6]. In the comparative research, these six methodologies were employed. The goal of this research is to develop a discovery framework that can predict when disease outbreaks will occur. The system employs SFS calculations, six classifiers, a cross-approval technique, and execution assessment metrics. A machine learning approach for planning the choice of a decision support network will make heart disease analysis more rational.

Using decision level fusion, U. Ahmed et al. suggested a machine learning-based diabetic decision assistance system. Using fuzzy logic, the proposed model incorporates two commonly utilized machine learning approaches. The accuracy of the suggested fuzzy decision system is 94.87, which is greater than that of other current systems. The authors can save countless lives with this diagnosing methodology [7]. Furthermore, the fatality rate from diabetes can be reduced if the condition is detected early and preventative actions are implemented.

S. Akter et al. suggested an approach that depends on seven deep learning algorithms for detecting CKD and identifying risk variables that are critical for early diagnosis and preventing the illness from progressing to the end stage. The research offers a comprehensive evaluation of deep learning algorithms' performance on CKD. The work contributes to the corpus of knowledge in the following ways: it utilized a scientific data processing technique to locate missing values in the CKD dataset. To fill in the missing values, the authors used linear regression and logistic regression for numerical and categorical data, respectively [8]. Five state-of-the-art feature selection processes were utilized to select features, followed by a comparison of the performance of the seven algorithms to explain the utility of including versus not including feature selection application in deep learning, adding statistical significance to establish the outcome's reliability, and finally, four DL models simple RNN, Bi GRU, Bi LSTM, and GRU were used to predict CKD for the first time.

The current work by J. Chen et al. proved that a deep learning technique may be utilized to identify clinical parameters linked to the development of IDH during an HD session. They attempted to use a deep neural network to detect susceptible factors such as demographics, comorbidities, laboratory parameters, vascular access parameters, reference values of HD machines during an IDH event, dialyzer components, and medications. The goal was to determine the importance of specific factors in the development of fast IDH. Therefore, to prevent fast IDH, appropriate procedures were expected to be implemented in specific circumstances. This is the first attempt, to their knowledge, at using a DNN model with clinical factors to predict the presence of IDH during an HD session [9]. The authors expect this model to improve its accuracy in predicting IDH in the future by incorporating additional clinical samples and covariates into the studies.

After obtaining blood samples and other individual characteristics, V. K. Daliya predicts how the diabetes illness would proceed over a year. Unlike the evaluated publications, instead of predicting whether a person is diabetic or not based on blood glucose levels, insulin, and food, this research focuses on the evolution of diabetes illness over one year [10]. The Multivariable Linear Regression approach is subjected to an optimization procedure involving feature reduction and logarithmic transformation. Therefore, this approach can provide insight into the kind of precautions and good practices that should be adopted to slow the advancement of the disease over time.

The suggested model by A. H. Syed et al. combines weighted average ensemble-based learning methods to create a prediction model with greater sensitivity, ROC AUC, and PR AUC values that can readily differentiate MCI individuals from healthy participants. On the other hand, obtaining all of these indicators from a single subject is expensive and unfeasible in multiple marker investigations [11]. Thus, the cost-effectiveness of the present ensemble technique is its distinct value, since profiling of innovative combination of CSF protein may properly diagnose patients with early stages of AD. The authors developed a web-based live forecasting system depending on a new combination of CSF protein biomarkers, the first of its kind for predicting the early stages of Alzheimer's disease.

A retrieval-polished (RP) response generating model was introduced by L. Zhang et al. [12]. The primary concept behind this framework is to utilize the retrieved response to enhance the knowledge and fluency of the produced answer. RP is made up of three components: a prototype selector (PS), a generation-dependent polisher (GP), and a polished response filter (PRF) (PRF). To be more explicit, the authors create a PS to retrieve a contextually similar prototype. They then refer you to a doctor to get a polished response. The PRF is utilized to determine the ultimate response, either the polished response or the prototype, depending on a context-sensitive score. To create the GP, the authors combine contexts and prototypes using a variant of the encoder-decoder architecture.

III PROPOSED METHODOLOGY

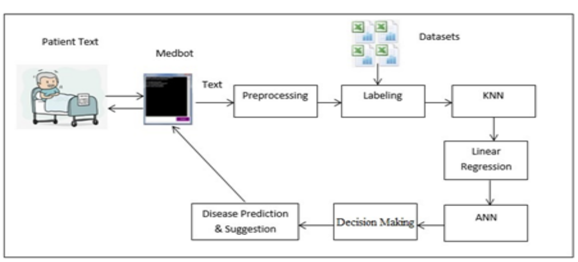

Figure 1: Proposed model System Overview

The proposed approach for a medical chatbot which uses machine learning to recognize and suggest diseases is detailed in the system overview shown in figure 1. The steps outlined below were used to accomplish the presented approach.

Step 1: Dataset collection, preprocessing and Labeling – The solution for the MedBot, which includes disease prediction and recommendation, involves three datasets: renal disease, Covid-19, and cardiovascular disease. The renal disease dataset may be obtained by going to this URL - https://www.kaggle.com/mansoordaku/ckdisease, the Covid-19 dataset from URL - https://data.gov.il/dataset/covid-19, and Cardiovascular Disease from URL - https://www.kaggle.com/ronitf/heart-disease-uci .

The datasets are retrieved and supplied as an input to the proposed methodology, which executes the labeling operation first by extracting the datasets in the format of a double-dimensional list. This procedure organizes the data and turns it into a labeled data. The interactive user interface, which is developed in the Python programming language and uses the Tkinter GUI toolkit to connect the Tk functionality to the Python code, also collects user input. This user interface notifies the user with the three illnesses' most prevalent symptoms. After the user chooses specific disease, the user can select disease-related parameters from the patient as an input and preprocesses that before sending it to the system for subsequent assessment.

Step 2: K Nearest Neighbor Clustering – This stage of the system receives the labeled list and user input as input to execute the clustering process. The renal disease, Covid-19, and cardiovascular disease databases are all in the format of a double-dimensional list. The K-Nearest Neighbors algorithm is used to divide input data into clusters that may be utilized to determine semantic groupings. The clusters are created by performing the following procedures.

Distance Evaluation – The Euclidean Distance is being used to calculate the distance between the user input and the specified characteristics of the input double dimension list. The row distance R_D is the distance calculated for the specified characteristics and affixed to the end of the row. This is done for each row in the list, and the relevant row distances are precisely inserted by using equation 1 below. These row distances are also evaluated for the average row distance, which is then suitably recorded.

$ED=\sqrt{(\sum(ATi-ATj))^2}$ ________ (1)

Where,

ED=Euclidian Distance

ATi=Attribute at index i
ATj= Attribute at index j

Centroid Estimation – As an input, the result from the previous section of distance measurement is offered. For centroid assessment, the list comprising the distances appended to the end of the rows is employed. The list is initially categorized into order of increasing of row distances in order to get the centroid. This sorted list is arbitrarily selected for data points. These data points were nothing but k-dimensional row distances. These row distances are therefore utilized to define the borders using the previously determined Average row distance.
The specified row distances from the data points, as well as the average row distance, are then utilized to generate the lowest and maximum values, which are formed by adding and subtracting both values. These limits will come in handy for forming the clusters in the subsequent stage.

Cluster Formation – In this stage, the k boundaries obtained in the preceding step are being used as a major factor in cluster formation. Relying on the boundaries established in the previous phase, the row distances in the double dimension list are examined. The row distances that comply by these bounds constitute the clusters, which are then saved as a cluster list and passed towards the next stage of the system.

Step 3: Linear Regression – The linear regression process calculates the relationship between user input and the cluster characteristics created in the previous stage. These data are used as input for the regression analysis in this stage, which employs Linear Regression. The regression analysis using linear regression detects and measures the differences between two separate variables. The two lists are x [] and y [], with x [] being the independent list and y [] being the dependent list. Equation 1 below shows the equation for the same.

Y=Mx + B ________________________ (2)

The slope provided as m and the magnitude of the intercept provided as b are undetermined in the equation above, hence the regression is assessed using that equation. These parameters are achieved by evaluating equations 3 and 4 as shown below. The user characteristics are represented by the values of x [], while the cluster characteristics are represented by the value of Y []. To get the needed values of m and b, these numbers are supplied to the equation.

$$M = \frac{\mathrm{N}\sum(xy) - \sum x \sum y}{\mathrm{N}\sum(x^2) - (\sum x)^2} \quad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_} \ (3)$$

$$B = \frac{\sum y - M\sum x}{\mathrm{N}} \quad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_} \ (4)$$

Where:
x = Independent variable
y = Dependent variable
M = Slope or Gradient
B = the Y Intercept
N= Size of the array
Y=Intercept value

The quantities of m and b obtained from the previous computations are then utilized in equation 2 to obtain the values of the dependent variables. In equation 2, an independent value from X[] is employed, and the results measured are summed to get the median regression value for every row in the cluster. The regression of the quantities of x [] and y[] allows for a better comprehension of the two variables' connection. These are the regression results from the regression analysis that are then compiled into a list and sent on to the following phase as input.

Step 4: Artificial Neural Network – The regression data analysis are initially categorized in decreasing order, and the first three clusters obtained from this technique are presented as input to this phase of the algorithm. These clusters are used by the Artificial Neural Network for output layer approximation via hidden layer assessment. To accomplish this, the 7 hidden layers are given a set of random weights depending on the number of input characteristics, as well as a bias weight for every layer. Equations 5 and 6 provide a quantitative description of the same.
This Artificial Neural Network receives the top three clusters with the highest similarity with the user input as input. The Artificial Neural Network is responsible for determining the hidden layer values from time series data in order to obtain the output layer values. The input layers are given 21 random weights with a bias of one. Including the contrast divergence and regular averaging of the input, the sigmoid activation function is applied. The equations 3 and 4 below show how this works mathematically.

$$T = \left(\sum_{k=0}^{n} AT * W\right) + B \text{\_\_\_\_} \ (5)$$

$$H_{LV} = 2\left(\frac{1}{(1+\exp(-T))} \ X \ 2T\right) - 1 \text{\_\_\_\_\_\_\_} \ (6)$$

Where,
n- Number of attributes

A_T- Attribute Values
W- Random Weight
B- Bias Weight
H_{LV} – Hidden Layer Value

The output layer of the Artificial Neural Network is determined by using the sigmoid activation function. After that, the results are saved in the format of a probability list. The highest probabilities are at the top of the list since the list is arranged in decreasing order. This is given to the next phase for the process of classification.

Step 5: Decision Making – This phase of the approach took into account the probability scores obtained via the usage of Artificial Neural Network deployment. The decision making approach is used in this stage to classify the probability values. This categorization is accomplished through the application of if-then rules that properly select the right output for illness diagnosis and recommendation creation. The categorization is also beneficial in reducing the amount of false positives that may enter the system. The system's recommendations are utilized to populate the MedBot user interface.

IV RESULTS AND DISCUSSIONS

The proposed machine learning approach for disease diagnosis and suggestion in a medical chatbot was implemented in the Python programming language. The Spyder IDE was used to ensure that the technique would be implemented effectively. The graphical user interface was created using the Tkinter toolkit, which was then linked to the Python code. The system was implemented on a developer computer with an Intel Core i5 CPU, 8GB of RAM, and 500 GB of hard disk space.

To assess the performance of the provided system, an experimental inspection of the technique is required. This characterizes the proportion of error the strategy achieves as well as whether the machine learning approaches have been correctly integrated in the system. The Mean Reciprocal Ratio is being used for purposes of analysis. As an outcome, the system generates a suggestion that must be reviewed for appropriateness by a person. This is because a human being is amongst the most accurate indicators of whether or not the recommended advice was suitable.

As a response, a group of ten users has been allocated to test the process by giving various input. For each of the three disorders, this is done repeatedly. Covid-19, Renal Disease, and Cardiovascular Disease Participants have assessed the results in the format of suggestions to determine their correctness.

The system's recommendations are rated by the participants. The suggestions are ranked from 1 to 6, with 1 being the best and 6 being the worst. The system is given various ranks, and an equivalent reciprocal of these ranks is calculated. This is accomplished by transforming the rank 2 to 1/2 and the rank 3 to 1/3, as well as the rank 6 to 0. The results are listed in table 1 below.

Pariticpant	Renal Disease	Cardiovascular Disease	Covid-19
1	1	0.5	1
2	0.5	0.5	1
3	1	1	1
4	1	0.5	1
5	1	1	1
6	0.5	1	1
7	1	1	0.5
8	0.5	1	1
9	0.5	0.5	0.5
10	1	0.5	1
MMR	0.8	0.75	0.9

Table 1: Participant Mean Reciprocal Ratio (MRR) for 3 different diseases

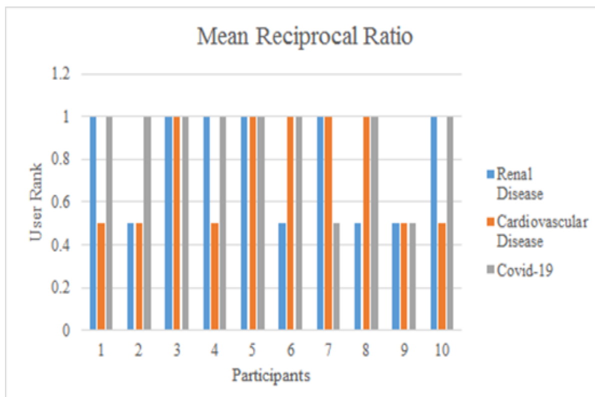

Figure 2: Graphical Representation of the MRR Values

The compiled results are used to create a graphical representation, which can be seen in figure 2 below. Examining the chart, it is clear that the system's proposal for Renal Disease, Cardiovascular Disease, and Covid-19 is extremely accurate, as it achieves MRRs of 0.80, 0.75, and 0.9, correspondingly, with an average MRR of 0.81. With the application of machine learning methodologies, the MedBot system is indeed very accurate and may be highly useful in giving much-needed aid to users by delivering important suggestions dependent on their symptoms.

V CONCLUSION AND FUTURESCOPE

The adoption of a medical chatbot that delivers a disease prognosis and effective recommendation to the particular patient has accomplished the strategy of functional enhancement in the healthcare perspective. The system accepts the patient's symptoms as input. These symptoms are efficiently preprocessed to offer a light weight query that may be quickly processed to arrive at a diagnosis. The query is then labeled using a collection of datasets comprising illnesses and symptoms. This labeling helps the system to appropriately reduce down the disease choices. The labelled text is then given to the K Nearest Neighbors in order to get accurate clusters. The resulting clusters are used in the following stage of the technique, which uses Linear Regression to retrieve the regression of the supplied symptoms. The regression list is then used in the next stage, which involves evaluating the ailment and providing predictions using Artificial Neural Networks. The Disease Prediction and Suggestion is provided by ANN, which is then successfully categorized utilizing the Decision-making Module's if-then rules. The user is then presented with the conclusions via a graphical user interface. Extensive investigation yielded positive results, allowing the technique to be adequately appraised for its functionality.

This approach can be improved in the future to be incorporated as a mobile application for patients and clinicians to operate easily with improved accessibility.

REFERENCE

[1] X. Ren, G. Spina, S. De Vries, A. Bijkerk, B. Faber and A. Geraedts, "Understanding Physician's Experience with Conversational Interfaces During Occupational Health Consultation," in IEEE Access, vol. 8, pp. 119158-119169, 2020, DOI: 10.1109/ACCESS.2020.3005733.

[2] B. Zhang, Z. Wei, J. Ren, Y. Cheng, and Z. Zheng, "An Empirical Study on Predicting Blood Pressure Using Classification and Regression Trees," in IEEE Access, vol. 6, pp. 21758-21768, 2018, DOI: 10.1109/ACCESS.2017.2787980.

[3] G. Mao, J. Su, S. Yu, and D. Luo, "Multi-Turn Response Selection for Chatbots with Hierarchical Aggregation Network of Multi-Representation," in IEEE Access, vol. 7, pp. 111736-111745, 2019, DOI: 10.1109/ACCESS.2019.2934149.

[4] M. Polignano, F. Narducci, A. Iovine, C. Musto, M. De Gemmis and G. Semeraro, "Health Assistant Bot: A Personal Health Assistant for the Italian Language," in IEEE Access, vol. 8, pp. 107479-107497, 2020, DOI: 10.1109/ACCESS. 2020. 3000815.

[5] R. Oruche et al., "Evidence-Based Recommender System for a COVID-19 Publication Analytics Service," in IEEE Access, vol. 9, pp. 79400-79415, 2021, DOI: 10.1109/ACCESS.2021.3083583.

[6] G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima and S. M. H. Akhter, "Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique with and Without Sequential Feature Selection," in IEEE Access, vol. 10, pp. 23808-23828, 2022, DOI: 10.1109/ACCESS.2022.3153047.

[7] U. Ahmed et al., "Prediction of Diabetes Empowered with Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, DOI: 10.1109/ACCESS.2022.3142097.

[8] S. Akter et al., "Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease," in IEEE Access, vol. 9, pp. 165184-165206, 2021, DOI: 10.1109/ACCESS. 2021.3129491.

[9] J. Chen, K. Wu, S. Moi, L. Chuang, and C. Yang, "Deep Learning for Intradialytic Hypotension Prediction in Hemodialysis Patients," in IEEE Access, vol. 8, pp. 82382-82390, 2020, DOI: 10.1109/ACCESS.2020.2988993.

[10] V. K. Daliya, T. K. Ramesh, and S. -B. Ko, "An Optimised Multivariable Regression Model for Predictive Analysis of Diabetic Disease Progression," in IEEE Access, vol. 9, pp. 99768-99780, 2021, DOI: 10.1109/ACCESS.2021. 309 6139.

[11] A. H. Syed, T. Khan, A. Hassan, N. A. Alromema, M. Binsawad, and A. O. Alsayed, "An Ensemble-Learning Based Application to Predict the Earlier Stages of Alzheimer's Disease (AD)," in IEEE Access, vol. 8, pp. 222126-222143, 2020, DOI: 10.1109/ACCESS.2020.3043715.

[12] L. Zhang, Y. Yang, J. Zhou, C. Chen and L. He, "Retrieval-Polished Response Generation for Chatbot," in IEEE Access, vol. 8, pp. 123882-123890, 2020, DOI: 10.1109/ACCESS.2020. 300 4152.