Smart Health Guidance Using Machine Learning

Kalaivani V[1], Manyam Vinod[2], Parthipan S[3], Rohith S[4]

Assistant Professor, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[1]

BE, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[2,3,4]

Abstract: Hospitals are the most widely used means by which a sick person gets medical check-ups, disease diagnosis and treatment recommendation. People consider it as the most reliable means to check their health status. The proposed system is to create an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get a diagnosis. If one is not very serious and only wants to know about the kind of disease facing, this system is the cure for all ills. Here, the system allows users to share their symptoms and gives the predicted disease with suitable treatment. Once the user enters the symptoms, the data is classified from the dataset and finally the disease will be predicted. After prediction this system also suggests suitable treatment for the disease and also the list of doctors specialized in that field and users can take appointments using this system.

Keywords: Machine Learning, Random Forest Classifier, Django, Data Preprocessing

I. INTRODUCTION

The health industry has been growing a lot in the past few years. Machine learning techniques have gained a lot of importance in medical areas. It has been calculated that a care hospital may generate five terabytes of data in the year. In our day to day life we have a lot of other problems to deal with and we neglect our health problems. So in order to overcome such problems we have designed a user-friendly website which helps users to get diagnosed from their residence at any time. We also provide an option for booking an appointment with the doctor to discuss health related problems and get diagnosed properly. Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

The main objective of this project is to develop a web application that helps patients to predict disease from symptoms and provide some treatment suggestions for the predicted disease along with the specialized doctors suggestion. The user can also book an appointment with the specialized doctor if needed.

The main problem here is that more time is taken to predict the disease corresponding to the problems. This has to be overcome. A general solution has to be developed which will provide facilities to resolve the health issues in a faster and more efficient way. The user has to search for the best doctor which consumes a lot of time and then book an appointment, so there is a lot of manual work for the user. This should be automated.

II. OVERVIEW

The dataset is taken from kaggle that consists of 132 symptoms, the permutations of which leads to 41 diseases. Based on the 4920 records of patients, we aim to develop a prediction model using Random forest classifier algorithm that takes in the symptoms from the user and predicts the disease user is more likely to have.

The flow of developing prediction model for predicting disease can be given as follows:

Input Symptoms

⬇

Data Preprocessing

⬇

Random Forest

⬇

Output Disease
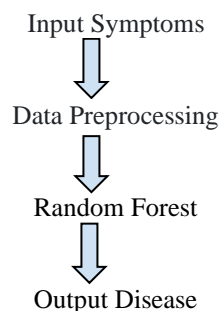
III. LITERATURE SURVEY

As per a description

[1] Sneha Grampurohit et al., proposed the **Disease Prediction using Machine Learning Algorithms.** This automated system can discover and extract hidden knowledge associated with the diseases from a historical(diseases-symptoms) database according to the rule set of the respective algorithms. It is developed to immensely help to solve the health-related issues by assisting the physicians to predict and diagnose diseases at an early stage. A Sample data of 4920 patients' records diagnosed with 41 diseases was selected for analysis. A dependent variable was composed of 41 diseases. 95 of 132 independent variables(symptoms) closely related to diseases were selected and optimized. This research work carried out demonstrates the disease prediction system developed using Machine learning algorithms such as Decision Tree classifier, Random forest classifier, and Naïve Bayes classifier. The paper presents the comparative study of the results of the above algorithms used. The dataset consists of 132 symptoms, the combination or permutations of which leads to 41 diseases. Based on the 4920 records of patients, it aims to develop a prediction model that takes in the symptoms from the user and predicts the disease he is more likely to have.

[2] Dahiwade, Dhiraj et al., proposed the Designing disease prediction model using machine learning approach.This proposed system is used for general disease prediction based on symptoms of the patient. For disease prediction, K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithms are used for accurate prediction of disease. For disease prediction the required disease symptoms dataset is used. In this general disease prediction the living habits of a person and checkup information consider for the accurate prediction. The accuracy of general disease prediction by using CNN is 84.5% which is more than KNN algorithm. And the time and the memory requirement is also more in KNN than CNN. After general disease prediction, this system is able to give the risk associated with general disease which is lower risk of general disease or higher.

[3] A. N. V. K. Swarupa et al., proposed the Disease Prediction: Smart Disease Prediction System using Random Forest Algorithm.This paper illustrates a disease prediction system constructed using the Random Forest Machine Learning algorithm. Experiments were conducted with a standard symptoms dataset, and the model achieved 95 % classification accuracy. Machine learning and the Python programming language with the Tkinter Interface were used to create the disease prediction using Random Forest.

[4] T. N. Deepak and M. M, has proposed the Prediction of Health Problems and Recommendation System Using Machine Learning and IoT.This paper proposes a Graphics user interface-based disease prediction system that utilizes seven classification Machine Learning algorithms to detect and forecast the problem based on symptoms. It also suggests alternate treatments and provides detailed information on the predicted disease. It also uses NodeMCU and ThinkSpeak to monitor the heart and temperature values over the internet, and it records all of the information in a real-time database that a doctor or user can access in the future.

IV. EXISTING SYSTEM

With the evolution of the algorithms in data mining. The prediction process is changing in terms of speed with the use of data mining techniques and new algorithms. But the existing systems lack in terms of speed and efficiency due to implementation of techniques with high time complexity and implementation of primitive algorithms. The existing systems only predicts the disease based on symptoms with some machine learning algorithms and won't provide any treatment or doctor suggestions and doctor appointments facility.

V. PROPOSED SYSTEM

The proposed system predicts the disease based on the symptoms given by the user using Random Forest Algorithm. Along with the disease prediction the proposed system also provides treatment and specialized doctors suggestions for the predicted disease. The patient can avail appointment of the required doctor using this system. The web application is developed using Django framework of python.

Advantages of Proposed system:
● User can predict disease using symptoms
● Treatment and doctor suggestion
● Doctor Appointment.

VI. ARCHITECTURE DESIGN AND MODULES

SYSTEM MODULES

There are five modules used in this project
1. Login module
2. Disease prediction module
3. Treatment Suggestion module
4. Doctor Suggestion module
5. Appointment module

Login module: In this module the patient can login to the application if already a valid user and holds credentials or can register if new to the application.

Disease prediction module: In this module the patient enters the symptoms and gets the predicted disease from the trained model based on the symptoms entered.

Treatment Suggestion module: In this module some treatment tips and suggestions are given to the patient for the disease predicted.

Doctor Suggestion module: In this module the specialized doctors list for the predicted disease is displayed along with their contact details.

Appointment module: In this module the patient can avail appointment with the specialized doctor.
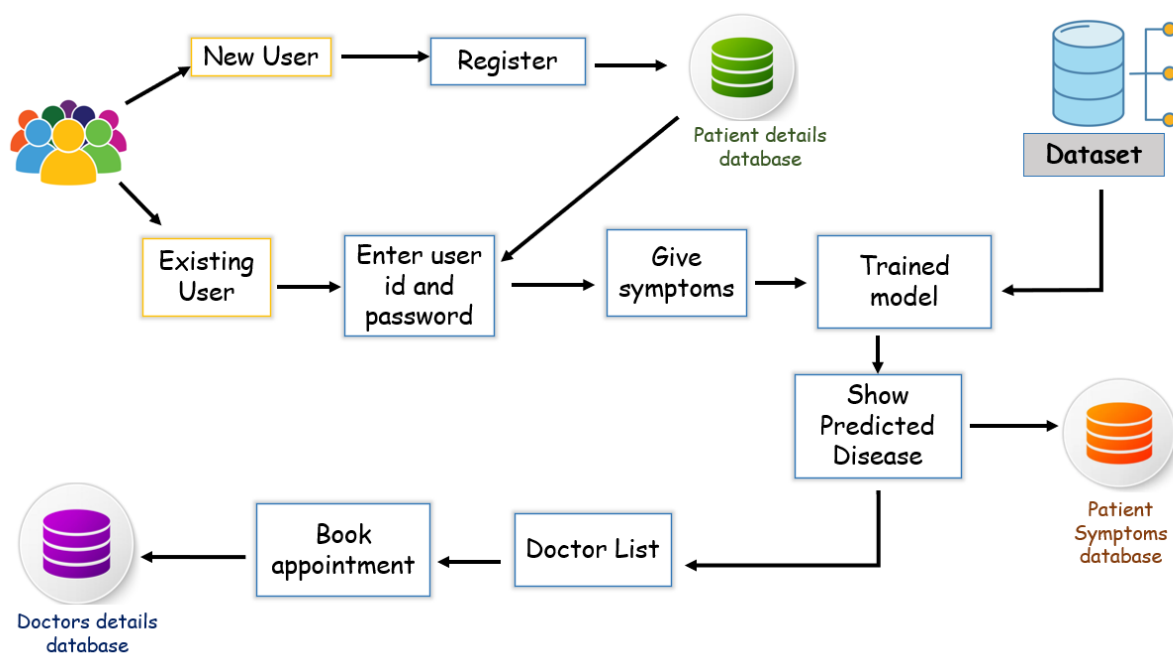

Figure 1: Architecture design

VII. METHODOLOGY

The disease prediction system is implemented using the machine learning algorithm Random forest classifier. The description and working of the algorithms are given below.

RANDOM FOREST ALGORITHM

A **Random Forest Algorithmic** program consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The random forest algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of overfitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

Random forest works as follows:

Step 1: Selects k symptoms from dataset (medical record) with a total of m symptoms randomly (where k<<m)Then, it builds a decision tree from those k symptoms.

Step 2: Repeats n times so that we have n decision trees built from different random combinations of k symptoms (or a different random sample of the data, called bootstrap sample).

Step 3: Takes each of the n-built decision trees and passes a random variable to predict the Disease. Stores the predicted Disease, so that we have a total of n Diseases predicted from n Decision trees.

Step 4: Calculates the votes for each predicted Disease and takes the mode (most frequent Disease predicted) as the final prediction from the random forest algorithm.

Performance of Algorithm on test data:

After training the model using random forest classifier algorithm, the model was tested on 41 new diseases. The accuracy score obtained for the test data is 97.6%
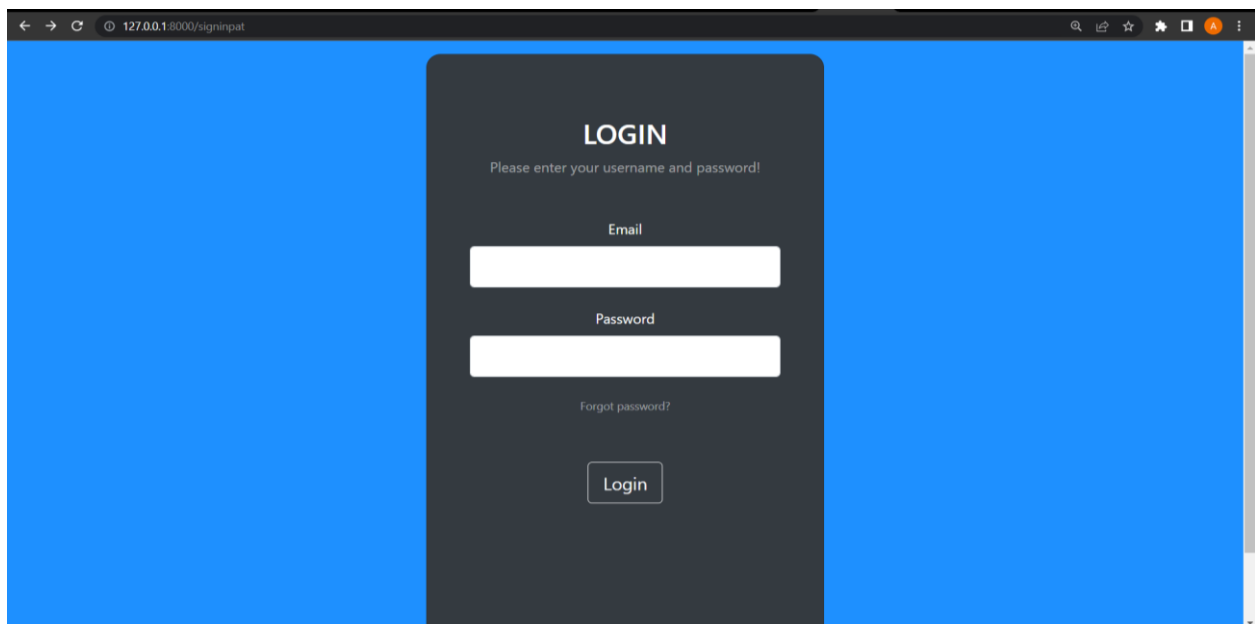
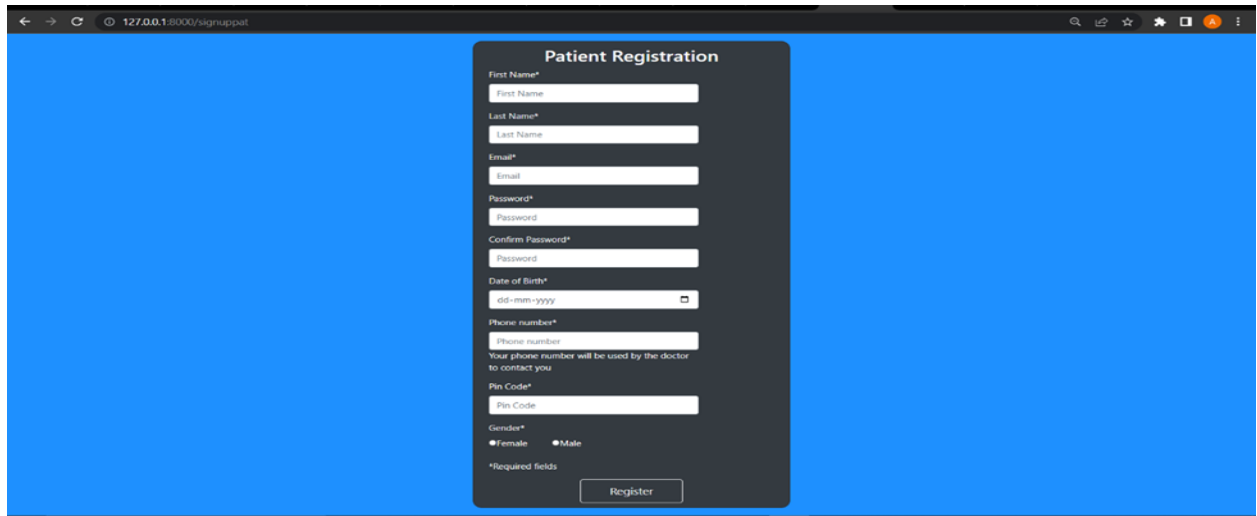## VIII. RESULTS

Figure 2: Login Page
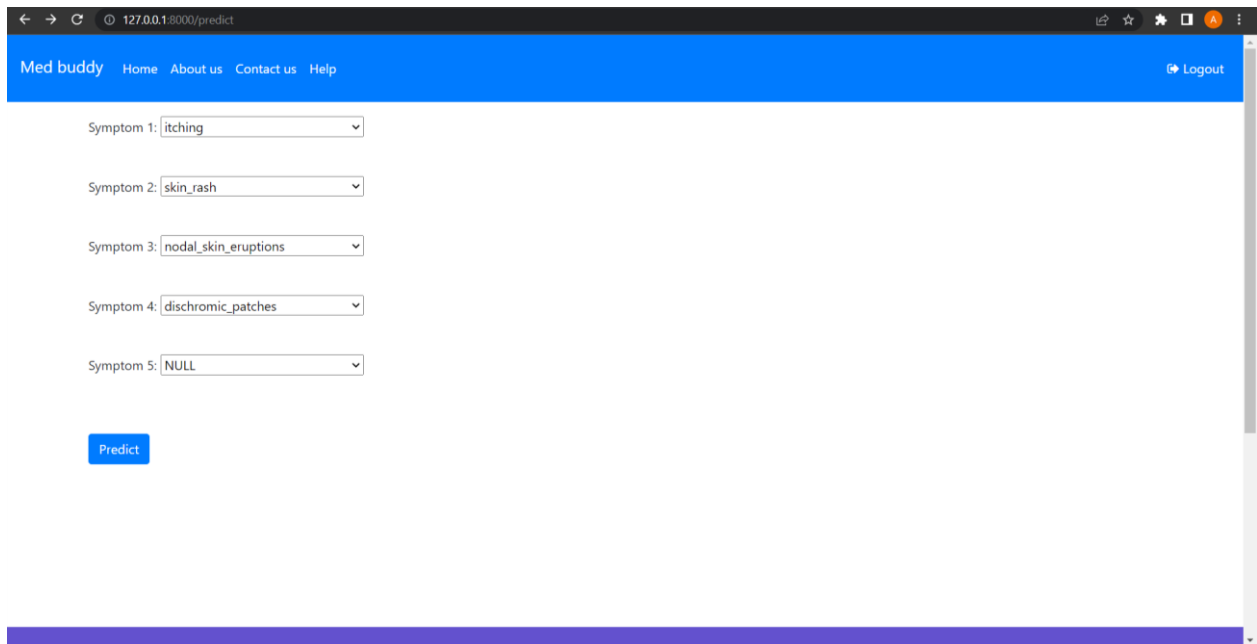
Figure 3: Signup Page

Figure 4: Entering Symptoms

The GUI created here takes 5 symptoms in which 3 symptoms are mandatory. The user can choose the symptoms from the list of symptoms that appears in drop down.
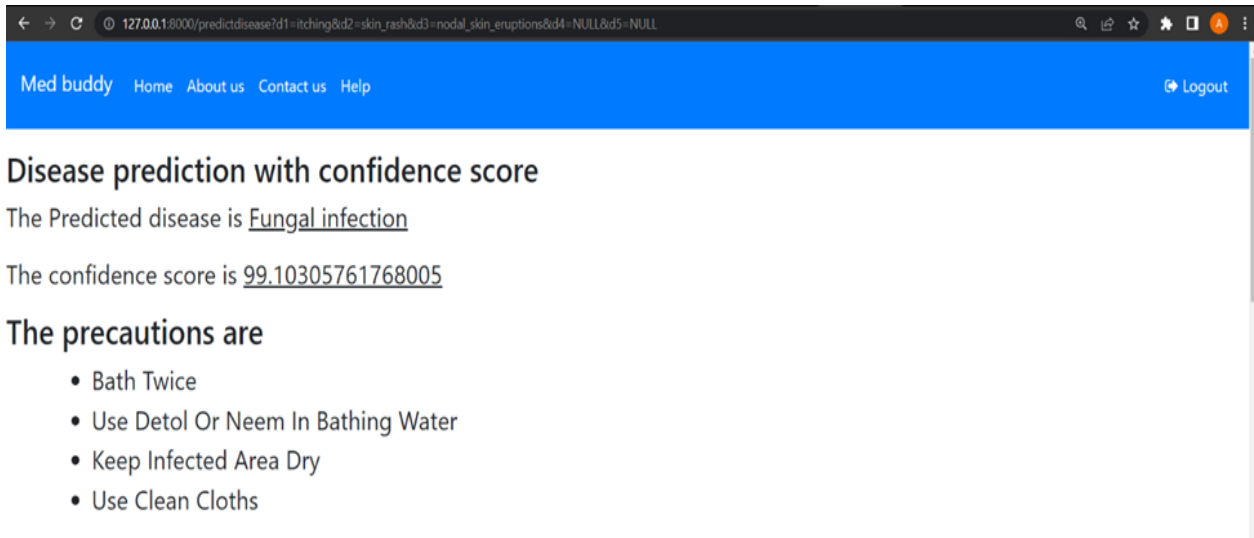
Figure 5: Predicted Disease And treatment Suggestion

After entering the symptoms the model built predicts the disease and the disease along with its confidence score and precautions are displayed.
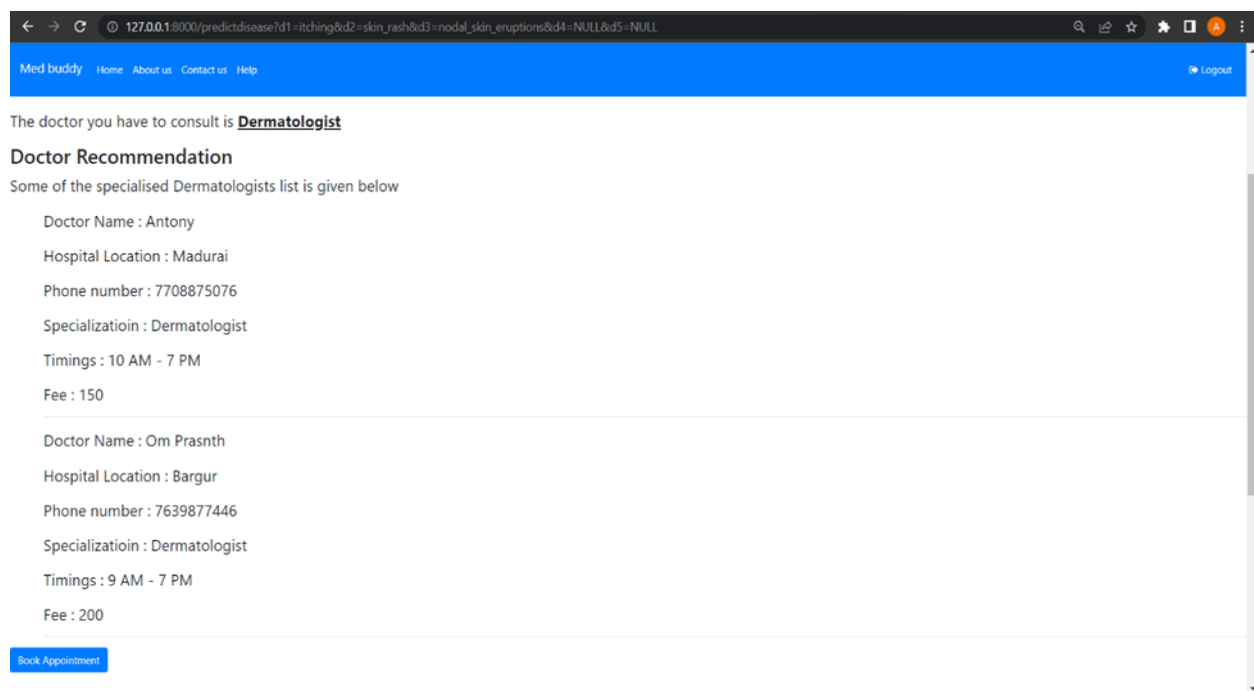

Figure 6: Doctor Suggestion

Along with the treatment suggesstion this system also suggests the specialised doctors for the predicted disease.

Figure 7: Booking Appointment

The patient can also book the appointment with the required doctor using this system

IX . CONCLUSION

This paper presents the machine learning model built using a Random Forest Classifier for predicting the disease based on the symptoms. This system involves fundamental parts like basic login, entering symptoms in the system and recommending medications, proposed and adjacent specialists. It takes the contribution of different symptoms from the patient and does the examination of entered symptoms and gives the name of the diseases.

REFERENCES

[1] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130.

[2] D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.

[3] A. N. V. K. Swarupa, V. H. Sree, S. Nookambika, Y. K. S. Kishore and U. R. Teja, "Disease Prediction: Smart Disease Prediction System using Random Forest Algorithm," 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT),2021,pp. 48-51, doi: 10.1109/ICISSGT52025.2021.00021.

[4] S. J P, T. N. Deepak and M. M, "Prediction of Health Problems and Recommendation System Using Machine Learning and IoT," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), 2021, pp. 1-8, doi: 10.1109/i-PACT52855.2021.9696622.

[5] Manzoor, M., Maqbool, M., Sarwar, A., Khan, M., llah, I.- u, Mir, H., Batool, B., Bangash, S. A., & Muthanna, F. M. (2022). Adults with Celiac Disease: Histopathological and Immunohistochemical Analysis of Small Intestinal Biopsies. Pakistan BioMedical Journal, 5(1), 249–252.

[6] M. Raja and G. G. L. Priya, ―Conceptual origins, technological advancements, and impacts of using Virtual Reality technology in education,‖ Webology, vol. 18, no. 2, pp. 116–134, 2021.

[7] Manzoor, M., Amin, A., Hussain, S., Zia, R., Sarwar, A., Shabir, M. M., llah, I.- u, Mir, H., Bangash, S. A., & Muthanna, F. (2022). Correlation Between Histopathological Findings, CD4 Counts, and Treponema Quantity in Microscopic Sections and Secondary Syphilis in HIV Positive Individuals. Pakistan BioMedical Journal, 5(1), 253–256.

[8] S. Venkatasubramanian, D. A. Suhasini, and D. C.Vennila, ―An Energy Efficient Clustering Algorithm in Mobile Adhoc Network Using Ticket Id Based Clustering Manager,International Journal of Computer Science and Network Security, vol. 21, no. 7, pp. 341–349, Jul. 2021.

[9] Venkatasubramanian, S., Suhasini, A. and Vennila, C., ―An Efficient Route Optimization Using Ticket-ID Based Routing Management System (T-ID BRM)‖. Wireless Personal Communications, pp.1-20, 2021

[10] Muthanna, F. M., Samad, A., Ibrahim, H. K., Al-Awkally, N.-A. M., & Sabir, S. (2022). Cancer related anaemia (CRA): An overview of approach and treatment. International Journal of Health Sciences, 6(S2), 2552–2558.

[11] J. Gao, L. Tian, J. Wang, Y. Chen, B. Song and X. Hu, "Similar Disease Prediction With Heterogeneous Disease Information Networks," in IEEE Transactions on Nanobioscience, vol. 19, no. 3, pp. 571-578, July 2020, doi: 10.1109/TNB.2020.2994983.

[12] Muthanna, F. M. S., Hassan, B. A. R., Karuppannan, M., & Mohammed, A. H. (2021). Evaluation of the impact of anemia on quality of life among breast cancer patients undergoing chemotherapy in Malaysia. Journal of Pharmaceutical Health Services Research, 12(2), 310- 312.

[13] Iqbal, M. S., Muthanna, F., Kassab, Y. W., Hassali, M. A., Al-Saikhan, F. I., Iqbal, M. Z., Haseeb, A., Ahmed, M., Khan, S. U., Naqvi, A. A., Islam, M. A., & Ali, M. (2020). Determinants of health-related quality of life among warfarin patients in Pakistan. PloS one, 15(6), e0234734.

[14] Muthanna, F. M., Karuppannan, M., Hassan, B. A. R., & Mohammed, A. H. (2020). Alabdullah, T. T. Y., Ahmed, E. R., & Ahmed, R. R. (2021). Organization features and profitability: Implications for a sample of Emerging Countries. Journal of Accounting and Business Education, 5(2), 43-52.

[15] Nor, M. I., Masron, T. A., &Alabdullah, T. T. Y. (2020). Macroeconomic fundamentals and the exchange rate volatility: empirical evidence from Somalia. SAGE Open, 10(1), 2158244019898841. Assessment of Risk Factors Associated with Anemia Severity among Breast Cancer Patients Undergoing Chemotherapy in Malaysia. Systematic Reviews in Pharmacy, 11(12), 2405-2411.