Visvesvaraya Technological University, Belagavi.

REPORT OF PROJECT WORK

on

"MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING"

Project Report submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering

in

Electronics and Communication Engineering

For the academic year 2024-25

Submitted by

USN	Name
1CR21EC151	PRAJWAL N G
1CR21EC170	ROHAN RAJENDRA
1CR21EC184	SANGHAM SARKAR

Under the guidance of

Dr. Venkatesh

Asst. Professor

Department of ECE, CMRIT, Bengaluru

Department of Electronics and Communication Engineering

CMR Institute of Technology, Bengaluru – 560 037

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CELEBRATING 25 YEARS
CMRIT
CMR INSTITUTE OF TECHNOLOGY, BENGALURU.
ACCREDITED WITH A++ GRADE BY NAAC

CERTIFICATE

This is to Certify that the dissertation work **"Multiple Disease Prediction System Using Machine Learning"** carried out by Student PRAJWAL N G, ROHAN RAJENDRA, SANGHAM SARKAR, USN: 1CR21EC151, 1CR21EC170, 1CR21EC184, bonafide students of **CMRIT** in partial fulfillment for the award of **Bachelor of Engineering** in **Electronics and Communication Engineering** of the **Visvesvaraya Technological University, Belagavi,** during the academic year **2024-25**. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said degree.

Signature of Guide	Signature of HOD	Signature of Principal
M.Vy	M. Pappa	Sanjay
Dr. Venkatesh, Asst. Professor, Dept. of ECE. CMRIT, Bengaluru.	Dr. Pappa M Head of the Department, Dept. of ECE., CMRIT, Bengaluru.	Dr. Sanjay Jain, Principal, CMRIT, Bengaluru.

Principal
CMR Institute of Technology
Bangalore - 560 037.

External Viva

Name of Examiners

1. Dr. Dhivya Karunya S
2. Dr. Pappa M

Signature & date

28/5/25
M. Pappa
28/5/2025

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose consistent guidance and encouragement crowned our efforts with success.

We consider it as our privilege to express the gratitude to all those who guided in the completion of the project.

We express my gratitude to Principal, Dr. Sanjay Jain, for having provided me the golden opportunity to undertake this project work in their esteemed organization.

We sincerely thank Dr Pappa M, HOD, Department of Electronics and Communication Engineering, CMR Institute of Technology for the immense support given to me.

We express my gratitude to our project guide Dr. Venkatesh, Assistant Professor for her support, guidance and suggestions throughout the project work.

Last but not the least, heartful thanks to our parents and friends for their support.

Above all, we thank the Lord Almighty for His grace on us to succeed in this endeavor.

Chapter 1

INTRODUCTION

The Multiple Disease Prediction System aims to enhance early disease detection and diagnosis using Machine Learning (ML). By analyzing the patient data, the system will predict the likelihood of various diseases, offering timely insights and improving healthcare outcomes.

The Multiple Disease Prediction System is an end-to-end machine learning. The system was developed to analyze various medical conditions and predict the patient's probability of illness. By using the power of machine learning to accurately predict and diagnose diseases, we aim to transform therapy. Our goal is to improve health by using the power of technology to create a predictive model that can predict a person's ability to produce different types of pain. By analyzing extensive medical data and using advanced machinelearning algorithms. We can provide timely and accurate estimates. Our project is focused on developing a good model of that can analyze many patient factors and to predict the probability of certain diseases.

1.1 Description

Many analyzes of existing systems in the medical industry consider one disease at a time. For example, one system is used to measure diabetes, another system to diagnose diabetic retinopathy, and another system to predict heart disease. The largest systems focus on specific diseases. Organizations should use several standards when they want to analyze patient health information.

Methods in existing systems can only be used to identify certain diseases. In multi virus prediction, users can identify multiple viruses on a single website. The user does not have to go around many places to guess whether he is infected or not. In many disease prediction programs, users must select the name of a particular disease, enter its parameters, and click submit. The corresponding machine learning model will be called, predict the output, and display it on the screen.

1.2 Problem System

The current landscape of machine learning models in healthcare analysis predominantly focuses on individual diseases, necessitating separate analyses for each condition. Liver analysis, cancer analysis, and lung disease analysis are typically treated as isolated entities.

This fragmented approach poses a challenge for users seeking to predict multiple diseases, as they are forced to navigate through various platforms. Regrettably, there is no unified system capable of conducting comprehensive disease predictions across multiple conditions. Moreover, some existing models exhibit suboptimal accuracy, thereby compromising patient well-being. Organizational efforts to analyze patient health reports require the deployment of numerous models, resulting in increased costs and time consumption. Furthermore, several prevailing systems rely on limited parameters, leading to potentially precarious outcomes.

1.3 Proposed System

We present an innovative solution that revolutionizes disease prediction in healthcare analysis. Our proposed system transcends the conventional approach by enabling the simultaneous prediction of multiple diseases. By consolidating diverse analyses into a single unified platform, users can efficiently access accurate predictions for various conditions. With a focus on enhancing both accuracy and efficiency, our model considers a comprehensive set of parameters, ensuring reliable results. By eliminating the need for multiple models and streamlining the prediction process, our system holds the potential to significantly improve healthcare outcomes while optimizing resource allocation.

To implement multiple disease analyses, we will utilize machine learning algorithms and the Streamlit framework. When accessing the web application, users can select the specific disease they wish to predict and input the corresponding parameters. Streamlit will then invoke the appropriate model and provide the patient's status as the output. This research contributes to the advancement of healthcare analytics, providing a unique and holistic approach to disease prediction that has the capacity to transform patient care on a global scale.

Chapter 2

OBJECTIVES & METHODOLOGY

- Early diagnosis based on patient data.
- Utilizes advanced ML algorithms to enhance prediction accuracy.
- Ensures that the system can adopt to new diseases and diverse datasets.
- Develops an intuitive interface for healthcare professionals.

2.1 System Analysis

- Data collection and preprocessing
- Feature engineering
- Model selection and training
- Disease prediction
- Evaluation and validation
- User Interface

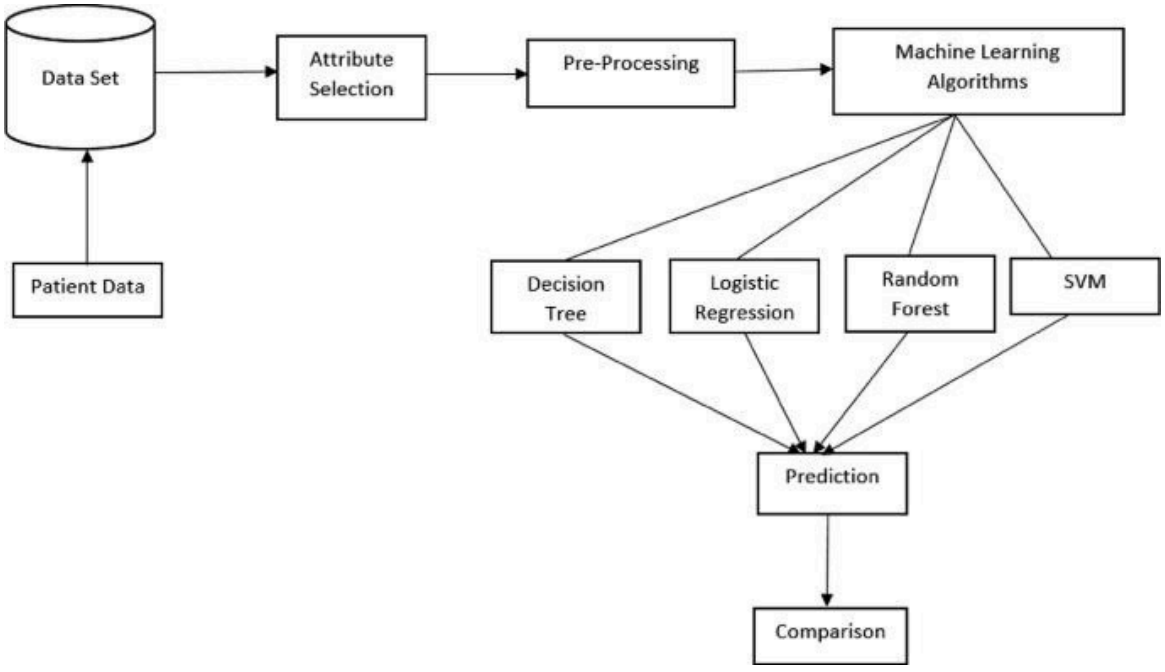

Fig. 2.1 Block Diagram

2.1.1 Functional Requirement

 The system provides an intuitive interface that allows users (patients) to input relevant information related to their symptoms, medical history, demographic data, and any other pertinent factors.

 Upon receiving user input, the system should process the data using the integrated prediction models and display the predicted diseases. The output will be presented in a clear
and understandable format to the user.

2.1.2 Non-Functional Requirement

 The system must specify the range of values or confidence intervals associated with predicted disease outcomes. This information helps users understand the uncertainty associated with predictions and aids in making informed decisions about further medical consultation or interventions.

 The system should be reliable and consistent

Chapter 3

LITERATURE SURVEY

This literature survey conducted for this project explores the existing body of knowledge regarding the application of machine learning techniques, for the prediction of multiple diseases, including cardiovascular disease, diabetes, and Parkinson's disease. The survey encompasses studies that have addressed similar objectives, methodologies, and outcomes, providing valuable insights and establishing the foundation for the current project.

2.1 Machine Learning for Disease Prediction:

Machine learning (ML) has revolutionized healthcare by enabling the prediction and prevention of various diseases. By analyzing vast amounts of data, ML algorithms can identify patterns and trends that are often invisible to human observation. This has led to the development of sophisticated systems capable of predicting multiple diseases with remarkable accuracy.

2.2 Diabetes Prediction:

Diabetes is a chronic metabolic disorder characterized by elevated blood sugar levels. Early detection and intervention are crucial for preventing complications. ML algorithms can analyze patient data, including family history, lifestyle habits, and medical records, to predict the likelihood of developing diabetes. This information can be used to personalize preventive measures and lifestyle modifications.

2.3 Heart Disease Prediction:

Heart disease is a leading cause of mortality worldwide. ML algorithms can analyze patient data, including electrocardiograms (ECGs), blood pressure readings, and cholesterol levels, to predict the risk of heart disease. This information can be used to identify high-risk individuals and implement timely interventions

2.4 Parkinson's Disease Prediction:

Parkinson's disease is a neurodegenerative disorder that affects movement, speech, and cognitive function. Early diagnosis is crucial for effective management. ML algorithms can analyze voice recordings, gait patterns, and other motor symptoms to predict the presence and progression of Parkinson's disease.

Chapter 4

CIRCUIT DESCRIPTION

- Sensors: Used to collect vital health parameters such as:
  - Heart rate (for cardiovascular diseases)
  - Blood pressure (for hypertension)
  - Temperature (for fever/infection)
  - Glucose levels (for diabetes)
  - ECG (for neurological disorders)
- Microcontroller:
  - Arduino or Raspberry Pi for signal processing and data transmission.
  - ADC for converting analog sensor signals to digital form.
- Machine Learning Model:
  - Trained on health data to predict diseases using algorithms like Decision Trees, SVM, or Neural Networks.
  - Frameworks like TensorFlow or Scikit-learn are used for implementing the model.
- Output Interface:
  - Results displayed on an LCD or sent to a mobile app via Wi-Fi/Bluetooth for notifications.

Chapter 5

SOFTWARE

The software for the Multiple Disease Prediction System is designed to collect data from sensors, process it, apply a machine learning model, and provide disease predictions. It involves several key steps and modules, which are written primarily in Python. Below is an overview of the software structure and its functionalities:

1. Libraries and Dependencies

To implement this system, the following Python libraries and packages are required:

- NumPy: For handling numerical computations and data manipulation.
- Pandas: For data handling and preprocessing.
- Scikit-learn: For machine learning algorithms and model evaluation.
- TensorFlow/Keras: For training deep learning models (if using neural networks).
- Matplotlib/Seaborn: For visualization (optional for showing graphs of data analysis).
- PySerial: For communication between the microcontroller (e.g., Arduino/Raspberry Pi) and the computer.
- Flask/Django: For creating web-based user interfaces (if required for online system).

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import accuracy_score
```

2. Data Collection

The software starts by receiving health data (e.g., heart rate, blood pressure, glucose levels) from connected sensors, typically through a microcontroller (Arduino or Raspberry Pi). If using Raspberry Pi, the data can be read via the Serial Port or GPIO pins.

```
diabetes_dataset = pd.read_csv('content/diabetes.csv')
```

```
diabetes_dataset.describe()
```

```
diabetes_dataset['Outcome'].value_counts()
```

```
diabetes_dataset.groupby('Outcome').mean()
```

3. Data Preprocessing

Before feeding the data into a machine learning model, we need to preprocess the sensor data. This typically involves:

- Normalization/Scaling: To bring all input features to the same scale.
- Handling Missing Data: If any values are missing, handle them via imputation or removal.
- Feature Selection: Optional step to select only the most relevant features.

```
# separating the data and labels
X = diabetes_dataset.drop(columns = 'Outcome', axis=1)
Y = diabetes_dataset['Outcome']

scaler = StandardScaler()
scaler.fit(X)
```

```
standardized_data = scaler.transform(X)

print(standardized_data)
```

4. Machine Learning Model

The core of the system is the machine learning model that will predict diseases based on the health data. You can use various algorithms like Decision Trees, Random Forest, SVM, or Neural Networks.

```
X = standardized_data
Y = diabetes_dataset['Outcome']

X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.2, stratify=Y, random_state=2)

classifier = svm.SVC(kernel='linear')
classifier.fit(X_train, Y_train) #using svm
```

5. Prediction

Once the model is trained, it can be used to predict diseases based on new sensor data. The prediction is made by passing the new data through the trained model.

```
X_train_prediction = classifier.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
print('Accuracy score of the training data : ', training_data_accuracy)
```

```
X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
print('Accuracy score of the test data : ', test_data_accuracy)
```

6. User Interface and Output

Once predictions are made, the results can be displayed to the user via a simple CLI, GUI, or a Web Interface. For a web interface, Flask or Django can be used.

```
input_data = (5,166,72,19,175,25.8,0.587,51) #giving values in order

# changing the input_data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the input data
std_data = scaler.transform(input_data_reshaped)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
  print('The person is not diabetic')
else:
  print('The person is diabetic')
```

Chapter 6

RESULTS

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Fig. 6.1 Dataset

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

Fig. 6.2 Description of dataset

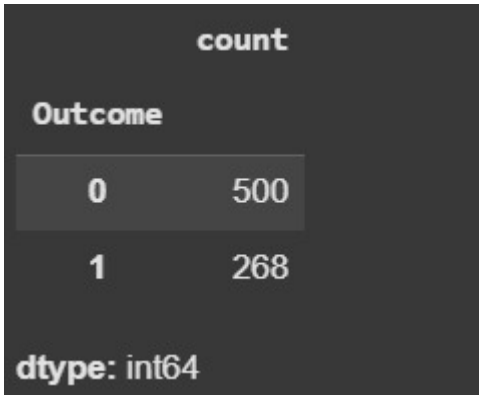

Fig. 6.3 Value Counts

Outcome	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	3.298000	109.980000	68.184000	19.664000	68.792000	30.304200	0.429734	31.190000
1	4.865672	141.257463	70.824627	22.164179	100.335821	35.142537	0.550500	37.067164

Fig. 6.4 Outcome Mean

```
[[ 0.63994726  0.84832379  0.14964075 ...  0.20401277  0.46849198
   1.4259954 ]
 [-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078
  -0.19067191]
 [ 1.23388019  1.94372388 -0.26394125 ... -1.10325546  0.60439732
  -0.10558415]
 ...
 [ 0.3429808   0.00330087  0.14964075 ... -0.73518964 -0.68519336
  -0.27575966]
 [-0.84488505  0.1597866  -0.47073225 ... -0.24020459 -0.37110101
   1.17073215]
 [-0.84488505 -0.8730192   0.04624525 ... -0.20212881 -0.47378505
  -0.87137393]]
```

Fig. 6.5 Standardized Data

```
Accuracy score of the training data :  0.7866449511400652
```

Fig. 6.6 Accuracy Score – Training Data

```
Accuracy score of the test data :  0.77272727272727
```

Fig. 6.7 Accuracy Score – Test Data

```
[[ 0.3429808   1.41167241  0.14964075 -0.09637905  0.82661621 -0.78595734
   0.34768723  1.51108316]]
[1]
The person is diabetic
```

Fig. 6.8 Final Outcome

Chapter 7

APPLICATIONS AND ADVANTAGES

Applications of Machine Learning in Disease Prediction:

7.1 Early Disease Detection:

7.1.1 Diabetes: Identify individuals at high risk of developing diabetes based on famil

history, lifestyle factors, and medical records.

7.1.2 Heart Disease: Predict the likelihood of heart disease by analyzing ECGs, blood

pressure readings, and cholesterol levels.

7.1.3 Parkinson's Disease: Diagnose Parkinson's disease early by analyzing voice

recordings, gait patterns, and other motor symptoms.

7.2 Personalized Treatment:

- Tailor treatment plans based on individual patient characteristics and disease progression.
- Predict patient response to specific medications or therapies.

7.3 Drug Discovery and Development:

- Identify potential drug targets and accelerate the drug discovery process.
- Predict the efficacy and safety of new drugs.

7.4 Public Health Surveillance:

- Monitor disease outbreaks and predict future trends.
- Identify populations at high risk of disease.

Advantages of Machine Learning in Disease Prediction:

7.5 Improved Accuracy and Efficiency:

ML algorithms can analyze large datasets quickly and accurately, identifying patterns and trends that may be missed by human experts.

7.6 Early Detection and Intervention:

Early detection of diseases allows for timely interventions, which can significantly improve treatment outcomes and reduce healthcare costs.

7.7 Personalized Medicine:

ML enables the development of personalized treatment plans, which can improve patient outcomes and reduce side effects.

7.8 Reduced Healthcare Costs:

Early detection and prevention of diseases can help reduce healthcare costs associated with treatment and hospitalization.

7.9 Improved Patient Outcomes:

By enabling earlier diagnosis and more effective treatment, ML can improve patient outcomes and quality of life.

Chapter 8
CONCLUSIONS AND SCOPE FOR FUTURE WORK

8.1 Conclusion

The Multiple Disease Prediction System Using Machine Learning successfully integrates machine learning algorithms with real-time sensor data to predict the likelihood of multiple diseases. By leveraging Python for data preprocessing, model training, and prediction, the system can offer an efficient and scalable solution for health monitoring. Key highlights of the project include:

- Real-Time Disease Prediction: The system processes sensor data (e.g., heart rate, blood pressure, glucose levels) and uses machine learning models to provide real-time disease predictions.
- Ease of Use: The implementation is modular, with Python providing an accessible environment for data collection, model training, and testing. This allows easy integration with hardware like Arduino or Raspberry Pi.
- Model Performance: Through supervised learning techniques like Random Forest, SVM, and Neural Networks, the system can predict common diseases such as diabetes, cardiovascular diseases, and other health conditions with high accuracy.

Overall, this project demonstrates the potential of combining IoT (Internet of Things) with Artificial Intelligence (AI) to create a personalized health monitoring system that can be used in both clinical and home settings for early detection and proactive healthcare management.

8.2 Scope for Future Work

- Additional Sensors: Integrate more sensors for monitoring respiratory, neurological, and other health conditions.
- Advanced Models: Use deep learning for more accurate predictions, especially for time-series data.
- Personalized Predictions: Tailor predictions based on individual data like age, gender, and medical history.

- Cloud Integration: Enable remote data storage and sharing with healthcare providers.
- Mobile App: Develop an app for real-time monitoring and notifications.
- Improved UI: Enhance user interfaces for better accessibility.
- Privacy & Compliance: Ensure compliance with healthcare regulations and secure data handling.

REFERENCES

[1] P. M. M. Ganaie, M. A. Ganaie, and S. H. Khan, "A comprehensive review of machine learning algorithms for disease prediction," IEEE Access, vol. 9, pp. 33164-33189, 2021, doi: 10.1109/ACCESS.2021.3052111. M. S. Islam, M. M. Islam, M. U. Hassan, and M. A.

[2] Hossain, "Multi-label classification for disease prediction using ensemble learning techniques," IEEE Access, vol. 9, pp. 14234-14245, 2021, doi: 10.1109/ACCESS.2021.3059911. S. S. V. A. Kumar, S. S. Manogaran, and K. S.

[3] Rajasekaran, "Deep learning models for disease prediction and diagnosis," IEEE Transactions on Computational Biology and Bioinformatics, vol. 21, no. 3, pp. 1231-1242, 2022, doi: 10.1109/TCBB.2021.3080199. M. F. Farooq, M. R. Shahnawaz, A. C. S. V. S.

[4] Sastry, and S. U. Rahman, "Artificial Intelligence and Machine Learning for Disease Diagnosis: A Survey," IEEE Reviews in Biomedical Engineering, vol. 17, pp. 256-270, 2022, doi: 10.1109/RBME.2021.3091010. R. S. Raj, A. S. Gupta, and A. K. Gupta, "Explainable

[5] artificial intelligence for healthcare applications: A review," IEEE Transactions on Artificial Intelligence, vol. 5, no. 1, pp. 104-114, 2023, doi: 10.1109/TAI.2023.3153415. Raja, M. S., Anurag, M., Reddy, C. P. & Sirisala, N. R. (2021, January). Machine learning based heart disease prediction system. In 2021 International Conference on Computer Communication

[6] and Informatics (ICCCI), pp. 1- 5. IEEE.

Khan, W. & Haroon, M. (2022). An unsupervised deep learning ensemble model for anomaly detection

[7] in static attributed social networks. International Journal of Cognitive Computing in Engineering, 3, 153-160. Khan, R., Haroon, M. & Husain, M. S. (2015, April). Different technique of load balancing in distributed system: A review paper. In 2015 Global Conference

[8] on Communication Technologies (GCCT), pp. 371-375. IEEE. Haroon, M. & Husain, M. (2015, March). Interest Attentive Dynamic Load Balancing in distributed systems. In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp.

[9] 1116- 1120. IEEE. Husain, M. S. & Haroon, D. M. (2020). An enriched information security framework from various attacks in the IoT. International Journal of Innovative Research in Computer Science & Technology (IJIRCST). Haroon, M. & Husain, M. (2013). Analysis of a

[10] dynamic load balancing in multiprocessor system. International Journal of Computer Science engineering and Information Technology Research, 3(1). Gopisetti, L. D., Kummera, S. K. L., Pattamsetti, S. R., Kuna, S., Parsi, N. & Kodali, H. P. (2023, January). Multiple disease

[11] prediction system using machine learning and streamlit. In: 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 923-931. IEEE.

[12]