

PROJECT_REPORT

by Dr.ramesh Sengodan

Submission date: 09-Jan-2025 09:52AM (UTC+0530)

Submission ID: 2561417272

File name: PROJECT_REPORT.docx (2.31M)

Word count: 9038

Character count: 53768

ABSTRACT

Artificial intelligence in medical diagnosis is transforming the world of healthcare, providing remarkable precision and efficiency.¹⁰ Artificial intelligence (AI) is a technology that allows computers and machines to replicate human learning, comprehension, problem solving, decision making, creativity, and autonomy. AI-enabled applications and devices can perceive and identify elements. They can comprehend and respond to human words. AI technologies, particularly in medical diagnostics, are changing the way diseases are identified, analyzed, and treated. AI can interpret large volumes of data quickly and effectively using ⁵⁴ machine learning and deep learning algorithms, providing healthcare providers with useful insights. These innovations not only improve diagnostic precision, but they also allow for early detection and individualized treatment methods. The advancement of AI in healthcare has proven transformational, particularly in the realm of medical diagnosis. AI was initially utilized mostly for tasks in administration, but its scope has grown enormously. AI and machine learning systems now evaluate large volumes of data rapidly and correctly, allowing healthcare providers to make better judgments. These technologies can process medical images, identify trends, and even forecast disease outcomes, transforming medical practice. In neglected and rural regions, access to high-quality healthcare is frequently confined, which often results in delayed diagnosis and suboptimal health outcomes. Existing ways of resolving this issue, such as telemedicine, have struggled to grow in parallel with growing demands for healthcare. This approach envisions an AI-powered system that can understand a vast amount of medical data, diagnose symptoms, and engage in discussions with patients to seek out about their concerns with their health. With the emergence of cutting-edge AI-driven technologies and the rising appeal of smart assistants like Google and Alexa, innovation in healthcare is entering an age of change. Consider a software that assesses symptoms, suggests diagnoses and treatment options. It may enable you to exercise self-care and track your progress. This seems like an ideal concept, particularly to

those areas with little healthcare access. While AI applications can help with initial health assessments, they shouldn't be used in place of expert medical guidance. In order to receive the best care, seek the advice of a doctor if symptoms intensify or continue. AI tools can serve as helpful reminders to put your health first and get help when you need it.

1 ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**,
Pro-VC, School of Engineering and ¹Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair**, School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Asif Mohammed**, Head of the Department, School of Computer Science Engineering & ¹Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr Ramesh Sengodan**, Professor and Reviewer ¹**Dr. Pamela Vinitha**, Professor, School of Computer Science Engineering & Information Science, Presidency University for his/her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman**, department Project Coordinators “Dr.Ramesh Sengodan” and Git hub coordinator ¹**Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Ms.Shreya Ravi Kumar

Ms.Sneha R

Ms.Neha R

7
LIST OF TABLES

Sl. No.	Table Name	Table Caption	Page No.
1	Table 1	Summary of Data Characteristics for CVD Prediction Models	10
2	Table 2	Engineered Features for Predictive Modeling	11
3	Table 3	Performance metrics of Machine Learning models for CVD Prediction	11

1 LIST OF FIGURES

Sl. No.	Figure Name	Caption	Page No.
1	Figure 1	Mobile CT system	9
2	Figure 2	Architecture	14
3	Figure 3	Gantt chart	18
4	Figure 4,5,6,7	Screenshots	26-27

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGENO.
	ABSTRACT	i
	ACKNOWLEDGMENT	ii
1.	INTRODUCTION	1
	1.1 AI and healthcare	1
	1.1.1 Use cases of AI in healthcare	1
	2.1 AI in disease diagnosis	2
	2.1.1 Applications used for providing diagnosis	2
	3.1 AI in the detection of rare diseases	3
2.	LITERATURE REVIEW	4
	2.1 Using AI, Diagnosis of Acute Diseases in Villages and Smaller Towns	4
	2.2 Diagnosis Of Acute Diseases In Villages And ¹² Smaller Towns Using AI	5
	2.3. Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter?	6
	2.4 Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19	7
	2.5 AI and Big Data: A New Paradigm for Decision Making in Healthcare	9
	2.6 Harnessing AI for Early Detection of Cardiovascular Diseases: Insights from Predictive Models Using Patient Data	10
	2.7 The Use of AI in Detecting Rare Diseases	11

3.	RESEARCH GAPS OF EXISTING METHODS	13
4.	PROPOSED METHODOLOGY	14
5.	OBJECTIVES	16
6.	SYSTEM DESIGN & IMPLEMENTATION	17
7.	TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)	18
8.	OUTCOMES	19
9.	RESULTS AND DISCUSSIONS	20
10.	REFERENCES	24
11.	APPENDIX-A PSUEDOCODE	25
12.	APPENDIX-B SCREENSHOTS	26
13.	APPENDIX-C ENCLOSURES	28

CHAPTER-1

INTRODUCTION

1.1 AI and healthcare

19

Artificial Intelligence (AI) technology allows computers and robots to mimic human learning, understanding, creativity, problem-solving, autonomy, and decision-making. It is revolutionizing healthcare by facilitating more rapid and precise diagnosis, individualized treatment procedures, and better patient outcomes. Artificial intelligence (AI)-powered solutions will assess enormous volumes of medical information, identify patterns, and forecast potential issues, assisting in the early diagnosis of conditions like diabetes and Parkinson's. Additionally, AI renders healthcare more efficient and accessible by boosting developments in telemedicine, remote monitoring, and drug discovery.

25

1.1.1 Use cases of AI in healthcare:

The use of AI in healthcare is widespread and significant, improving many facets of patient care and medical procedures. Medical imaging is one of the primary use cases, where AI algorithms evaluate scans to increase the precision of diagnosis for diseases like cancer and pneumonia. Another major area is personalized treatment, where AI assesses patient data to suggest customized treatments based on medical and genetic histories. Furthermore, AI is essential for managing Electronic Health Records (EHRs), identifying healthcare fraud, and enabling remote patient care via chatbots and virtual assistants. These tools facilitate proactive monitoring of chronic illnesses, increase patient engagement, and accelerate administrative procedures. By examining enormous datasets to find patterns and forecast results, AI also advances medical research, clinical decision-making, and pandemic preparedness. In the end, it improves the effectiveness of healthcare delivery by supporting medicines discovery, early disease identification, and clinical trial optimization. Healthcare professionals may concentrate more on patient care thanks to AI's ability to automate documentation, solve staffing issues, and improve mental health monitoring. The healthcare sector is utilizing AI to provide better, quicker, and individualized care, which is increasing patient outcomes and overall operational efficiencies.[11]

1

2.1 AI in disease diagnosis

46

Artificial intelligence (AI) has shown itself to be a useful tool for diagnosing illnesses, helping doctors analyze medical imaging such as CT, MRI, and X-rays for quicker and more precise results. Through the processing of patient information, symptoms, and medical history, AI can help with future diagnosis. As this field develops, artificial intelligence (AI) is anticipated to find patterns in massive datasets, allowing for the combining of genetic, lifestyle, and environmental data to diagnose complicated illnesses and the prediction and prevention of disease before symptoms appear. Although AI improves diagnostic procedures, it should be used in conjunction with medical practitioners' knowledge to improve patient care rather than taking their place. For AI to be developed and used in healthcare responsibly, ethical issues, data protection, and ongoing validation are crucial.

2.1.1 Applications used for providing diagnosis

With developments such as Quantum AI (QAI) and General AI (GAI), which provide increased processing capacity and real-time analysis of large medical datasets for more accurate and effective diagnoses, the future of AI-based medical diagnostics appears bright. While GAI, through programs like DeepQA, Watson, and DeepMind, enables pattern identification and data correlation to enhance medical outcomes, quantum optimization algorithms may help determine the best course of therapy. [12]

3.1 AI in the detection of rare diseases

AI assists in the integration and analysis of diverse data. These systems have been utilized successfully over the years for a variety of well-known application cases. The use of these methods to rare diseases (RDs) is also valuable. Faster and more accurate diagnosis would be helpful for the RDs, which are referred to as orphan illnesses. In order to find new instances, algorithms have been developed and are currently being utilized to develop networks and store data from people with uncommon diseases. AI is a crucial diagnostic instrument for RDs as it could assist in genetic analysis, picture recognition, and clinical

decision-making.

A mere five percent of the more than 7000 RDs that have been reported globally have a medical treatment. The difficulties faced by RDs can be addressed by utilizing AI technology' capacity to combine and evaluate data from many sources.

CHAPTER-2

LITERATURE SURVEY

2.1 Using AI, Diagnosis of Acute Diseases in Villages and Smaller Towns [3]

Proposed Scheme: There are a number of important steps in the systematic process of developing an AI-powered healthcare system. The system's objectives and scope must first be established, with an emphasis on certain issues such as disease detection, enhancing accessibility, or aiding medical professionals. Data collection, which includes compiling medical datasets while maintaining data privacy and regulatory compliance, is then essential. It is also necessary to describe the technological requirements, including scalability and AI model design. After choosing pertinent features from the data, the AI model is created, trained, and improved continuously. Essential elements like chat, login, and user pages are all part of a simple user interface that is made to make it easier for patients, healthcare providers, and the system to communicate. The system is then put through a rigorous testing and validation process to guarantee its accuracy in practical settings. System performance, user input, and data quality must all be constantly tracked for continuous improvements. Following continuous iterations to improve the AI model based on actual usage and performance data, the system's effect on patient outcomes and healthcare efficiency is finally evaluated.

A system's components, processes, and user interactions are all described in detail in the system design phase, which is a crucial stage in the development process. By covering both functional and technical elements, such as architecture, data structures, algorithms, and interfaces, it closes the gap between abstract concepts and implementation. System design provides a thorough roadmap that prioritizes work, maximizes resource allocation, and promotes effective development. This phase makes it possible for developers to plan ahead, predict problems, and come up with solutions, which guarantees a seamless transition from idea to reality and opens the door for a successful system or application.

Implementation details:

Intuitive Interface: Provide a user-friendly layout with intuitive navigation, eye-catching graphics, and thorough instructions.

Personalization: For user convenience, provide response regeneration, conversation management, and customized suggestions.

Medical care Suggestion: Offer the best available options for treatments while taking patient-specific characteristics, medication interactions, and best practices into account.

Easy access: Make sure it works with assistive technology and allows voice and text-based communication.

Patient Autonomy: Honor the rights and desires of the user without forcing judgments.

The outcomes of the research include increased productivity for healthcare providers by handling routine situations, reducing administrative work and data processing, improving diagnosis accuracy with accurate AI-driven insights, and improving time management.

Increased accessibility through remote help, financial savings through virtual consultations, patient empowerment through tailored health information, and timely resolution of common issues reduce wait times and exposure in medical facilities.

31

2.2 Diagnosis Of Acute Diseases In Villages And Smaller Towns Using AI [2]

Rural and underserved areas continue to face major barriers to accessing high-quality healthcare, which frequently leads to delayed diagnosis and worse health outcomes. The purpose of this research article is to examine how artificial intelligence (AI) can help address these inequities. In villages and small towns, it investigates how AI can improve the detection of acute diseases by analyzing previous research and literature. The results demonstrate how AI can revolutionize the delivery of accurate and easily accessible diagnostics to marginalized areas.

Methodology: A large dataset gathered from health information surveys was used to train the Support Vector Machines (SVM) and Decision Tree models used in the research technique. With a remarkable 91% accuracy rate, these models prove to be useful in the diagnosis of many illnesses. The system has a chatbot interface to enhance user engagement and guarantee accessibility in remote regions. The system's backend architecture uses the Flask framework to effortlessly combine chatbots and machine learning models, making it efficient and flexible. While the SVM model provides secondary analysis by combining other data, such as symptom duration, to produce severity scores, the Decision Tree Classifier delivers initial symptom-based predictions. In order to facilitate informed responses, information is loaded from CSV files during initialization, including symptom descriptions, severity rankings, and preventative measures. API endpoints that process user inputs, such as /predict and /svm1, provide comprehensive projections for the future descriptions of diseases, and suggestions. By allowing conversational symptom input and health assessments, the chatbot interface improves accessibility and makes the system easy to use,

particularly for users in rural areas. The predictive disease analysis, severity evaluation, thorough disease information, and physician consultation assistance are some of the system's primary features. When the app is first launched, the machine learning models are loaded to guarantee prompt and precise predictions. Predictions are enhanced by aggregated data from CSV files, and user inputs are processed smoothly through POST requests. Prediction findings and additional data are included in the structured JSON answers that the system produces. The SVM and Decision Tree models work well together for initial health evaluations, with an accuracy rate of 91%. Enhancing user engagement and establishing the system as a trustworthy, easily accessible resource for directing users toward the right healthcare steps are two benefits of integrating the chatbot interface.

¹² 2.3 Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter? [10]

This review was conducted using PubMed's Entrez Programming Utilities (API) to search the Medline database using a mix of keywords associated with rare diseases (RDs), congenital disorders of glycosylation (CDG), ⁴⁸ artificial intelligence (AI), machine learning (ML), and neural networks. Using Biopython libraries, a Python script made the search easier. The results were narrowed down to the top twenty pertinent papers using MeSH terms, titles, and abstracts. Three rounds of screening were used in the selection process: (1) four researchers screened the articles based on their titles and abstracts, (2) the full-text reviews of the articles that met the selection criteria took place in the second round, and (3) two independent researchers performed a final round of full-text analysis to guarantee consistent selection criteria for the included manuscripts. The following criteria were used for inclusion and exclusion: We only took into consideration English-language manuscripts ¹² that included the full title, abstract, and MeSH terms. Articles that addressed the use of certain AI algorithms (or sets of algorithms) to solve RD-related issues were included.

Reviews, although some were presented for context, were not included; only RDs with Orpha codes from the Orphanet categorization were taken into consideration. The included reviews' references were additionally checked to make sure no pertinent papers were missed. Table 1 lists the AI- and ML-based techniques that were used in the chosen articles.

2
Diagnosis of Acute Diseases in villages and smaller towns using

Table 1. List of available artificial intelligence (AI)- and machine learning (ML)-based methods used in rare diseases (RDs) organized by function.

General Function	Specific Function	Reference	Software/Platform/Algorithm	AI/ML Method	Disease(s)
Mutation Detection and Prediction	Predicts the pathogenicity/disease relevance of genetic variants	Airezaie et al.	CADD https://cadd.gs.washington.edu/	SVM	
			ClinPred https://sites.google.com/site/clinpred/	Ensemble classifier (RF and gradient boosting models)	Several RDs
		Yan et al.	CNVdigest https://github.com/yangx1016/CNVdigest	DNorm (conditional random fields, stochastic gradient descent, pairwise learning to rank)	Digeorge syndrome
			Fathmm-MKL http://fathmm.biocompute.org.uk/fathmmMKL.htm	SVM based on multiple kernel learning	
			GenoCanyon http://genocanyon.med.yale.edu/	Unsupervised statistical learning	
		Airezaie et al.	M-CAP http://bejerano.stanford.edu/mcap/	Gradient boosting trees	Several RDs
			MetaLR	Ensemble classifier	
			Meta-SVM	Meta-analytic SVM	
		Browne et al.	Meta-SNP http://snps.biofold.org/meta-snp/	RF	
			nsSNP Analyzer http://nsnpanalyzer.uthsc.edu/	RF	Mevalonic kinase deficiency
			PhD-SNP http://snps.biofold.org/phd-snp/phd-snp.html	SVM	

The diagnosis and prognosis of rare diseases (RDs) are being transformed by **artificial intelligence (AI)** and **machine learning (ML)** techniques that **improve** biochemical fingerprinting, phenotypic analysis, and mutation discovery. While SpliceAI and other deep learning frameworks evaluate cryptic splice mutations and noncoding variants, VEST and ClinPred employ machine learning (ML) to predict mutations that cause disease. Clinical data integration and phenotype-driven systems such as HANRD and eDiVA enhance variant prediction and disease gene prioritization. DeepGestalt and other imaging-based tools use facial recognition and neuroimaging to help diagnose RD, while biochemical techniques like infrared spectroscopy in conjunction with neural networks provide more affordable diagnostic options. AI is also improving prognostic modeling for diseases like synovial sarcoma, improving survival forecasts, and providing therapeutic advice for rare cancers and RDs.

9
2.4 Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19 [4]

4
There are three primary steps in the **imaging-based diagnosis** workflow for **COVID-19** using **thoracic CT** scans: image capture, pre-scan preparation, and disease diagnosis. As part of the pre-scan phase, technicians help patients position themselves in a certain way. The region from the apex to the base of the lung is covered by a single breath-hold CT scan during

image collection. Following the optimization of parameters according to the patient's body shape, the raw data is converted into pictures and sent to Picture Archiving and Communication Systems (PACS) for diagnosis.²⁹ By making imaging procedures safer, more precise, and more effective, artificial intelligence (AI) has completely transformed this conventional workflow. Specialized imaging platforms, lung and infection zone segmentation, clinical evaluations, and diagnostic tools are among the AI uses in COVID-19 imaging. Numerous commercial systems effectively incorporate AI to improve these procedures. During the COVID-19 pandemic, AI had a revolutionary effect on medical imaging, as evidenced by events like China's Medical Imaging Computing Seminar (MICS), which brought together academics and companies to exchange breakthroughs. This study tries to direct future research and applications while highlighting the vital role AI-powered medical imaging plays in the fight against COVID-19. It offers publicly accessible datasets to facilitate future developments and discusses intelligent imaging platforms and machine learning techniques for segmentation, diagnosis, and prognosis. The paper also discusses the difficulties and unresolved issues in the sector, providing useful information for radiologists and researchers to improve AI applications in medical imaging, especially when it comes to controlling pandemics like COVID-19.

Traditional imaging procedures have been revolutionized by AI-powered contactless imaging workflows, especially during the COVID-19 epidemic, as they minimize technician-patient contact and lower the danger of viral exposure. The intimate contact required for patient positioning in traditional procedures, including chest X-rays and CT scans, presents a serious risk of infection. With the use of sophisticated visual sensors like RGB, thermal cameras, and depth sensors, AI has made it possible to automate and contactless imaging procedures. These systems are capable of generating precise 3D patient models, identifying the anatomy of the patient, and estimating scanning parameters like scan range and ISO-centering. Access, efficiency, and safety have been further enhanced by mobile CT platforms with AI-based pre-scan and diagnosis tools, which separate personnel from patients and automate crucial processes like alignment and scanning. Furthermore, U-Net and its variations are examples of AI-powered segmentation approaches that have proven essential in locating lung regions and lesions in imaging data, enabling accurate diagnosis and therapy planning for COVID-19 and related illnesses. These developments show how AI is essential for improving imaging workflows and solving public health issues.

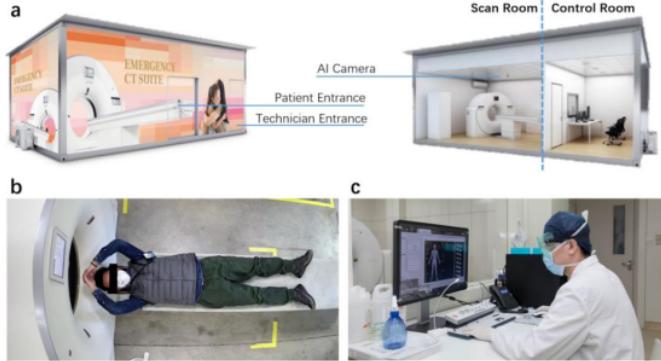


Fig. 1. (a) A mobile CT platform equipped with AI-empowered automated image acquisition workflow; (b) An example image captured by patient monitoring camera of CT system; (c) Positioning and scanning of patient operated remotely by a technician.

29

The patient is prompted to situate themselves on the patient bed both visually and audibly as soon as they enter the scan room. The patient can be observed through a window or through a live video feed from an AI camera positioned on the ceiling, and technicians can change the patient's posture as necessary. Using camera-captured photos, the patient positioning system creates a 3D posture and a completely reconstructed digital mesh of the patient after the technician or an AI-driven motion analysis algorithm has verified the patient's positioning.⁴ In order to provide control signals and optimum scanning settings for technician assessment, this 3D mesh is utilized to ascertain the scan range and the 3D centerline of the target body area.

The patient bed is automatically positioned in the ISO center and transferred into the CT scanner after approval. After that, the CT scans are processed and examined for screening and diagnosis.

6

2.5 AI and Big Data: A New Paradigm for Decision Making in Healthcare

The responsibilities of doctors and nurses could be drastically changed by the introduction of artificial intelligence (AI) into healthcare. Medical professionals will keep an eye on AI-driven choices and processes in the future to protect patients and reduce hazards like statistical inaccuracies and uneven treatment results. Acute care AI systems need to give thorough information about the variables affecting their choices in order to do this. Medical education must move from traditional decision-making to contemporary time-pattern

50

recognition in order to prepare doctors for AI oversight. Large amounts of varied and fast-moving data are

produced by the healthcare industry from sources such as lab tests, biological pictures, medical records, and real-time health monitoring. Big data approaches have enormous promise to improve disease identification and treatment as well as healthcare efficiency. The use of big data in health informatics is being investigated by researchers more and more, especially in the areas of complex disease management and diagnosis. [5]

11

2.6 Harnessing AI for Early Detection of Cardiovascular Diseases: Insights from Predictive Models Using Patient Data

Introduction: Every year, cardiovascular diseases (CVDs) claim around 18 million lives, making them a major global health concern. Early detection of these "silent killers" can often be prevented by current diagnostic techniques, which postpones essential therapies. Nonetheless, developments in machine learning (ML) and artificial intelligence (AI) offer chances for early diagnosis, which could revolutionize healthcare by accurately forecasting cardiovascular events.

Background and Motivation: A substantial amount of death worldwide is caused by CVDs, which are made worse by variables including aging populations and bad lifestyles. Conventional diagnostic techniques frequently only identify illnesses when they are somewhat advanced. Through the integration of many data sources, AI can improve risk assessment and aid in the early detection and management of CVDs. ⁵³

Methodology: Wearable technology, medical histories, and ECG readings were used to gather data. To increase predictive capability, key features were designed, and a number of machine learning models (ML) were created and verified, including random forests, support vector machines, and neural networks. Strong model performance was guaranteed by cross-validation and hyperparameter adjustment.

Data source	Number of Records	Number of Features	Feature Types
ECG Signal	50,000	15	Numerical (e.g. peak intervals)
Wearable Devices	60,000	10	Numerical (e.g. heart rate, activity levels)
Medical History	40,000	20	Categorical (e.g. family history, lifestyle factors)

Table 1: Summary of Data Characteristics for CVD Prediction Models

Data Type	Engineered Feature	Description
ECG Signal	Heart Rate Variability	Standard deviation of RR intervals
Wearable Devices	Activity Level Variance	Variance in step count over time
Medical History	Age-cholesterol Interaction	Interaction term between age and cholesterol

Table 2: Engineered Features for Predictive Modeling

Results: With a 92% accuracy rate, Gradient Boosting was the most successful model, followed by Support Vector Machines and Decision Trees. Age, blood pressure fluctuations, and cholesterol levels were identified as important predictors via feature significance analysis.

Gradient Boosting	92%	91%	90%	90.5%
Model	Accuracy	Precision	Recall	F1 score
Support Vector Machine	86%	84%	85%	84.5%
Decision Tree	89%	87%	88%	87.5%

Table 3: Performance metrics of Machine Learning models for CVD Prediction

By integrating a variety of data, AI has the potential to significantly improve early CVD prediction. In order to address ethical issues, test these models in practical situations, and broaden data sources for better healthcare outcomes, the findings support the need for more study. In order to lessen the burden of CVDs worldwide, proactive, tailored interventions powered by AI technology may be a part of the future of cardiovascular care. [6]

2.7 The Use of AI in Detecting Rare Diseases

This article examines how artificial intelligence (AI) can help with rare disease detection and treatment, which impact less than 1 in 2,000 people.

Difficulties with Diagnosis: Because rare diseases have a wide range of symptoms and a diagnostic process that might take five to thirty years, they frequently go undiagnosed. The lack of approved medicines for many diseases results in poor medical treatment.

AI's Role: Through the analysis of extensive datasets, symptoms, and genetic data, AI can improve diagnostic accuracy. The development of symptom-disease models with the aid of machine learning algorithms could speed up the diagnostic process.

Ethical challenges: The application of AI brings up ethical challenges, such as privacy concerns and biases in training datasets. AI has the potential to revolutionize the diagnosis of uncommon diseases, but its responsible and equitable application in healthcare requires addressing ethical issues.

Overall, the study highlights how AI has the ability to transform the detection and treatment of uncommon diseases while urging careful evaluation of the ethical ramifications.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

- Lower accuracy performance of the Support Vector Machine highlighted inverse complexity.
- The use of retrospective data and the biases present in electronic health records (EHRs) could affect how well these results generalize.
40
- AI models may inherit biases from training data, potentially leading to unequal treatment across different demographic groups. This makes ensuring fairness and transparency a challenge.
- Clinical deployment requires rigorous testing and validation in real-world settings to ensure reliability.
- Predictive models like Gradient Boosting can struggle with real-world accuracy due to overfitting or biased data.
- AI models in imaging require large, high-quality datasets, which can be challenging to obtain due to privacy and regulatory constraints.
- Reliance on AI can reduce physicians' independent decision-making, potentially leading to overdependence on technology.
- AI technologies can enhance efficiency and accuracy, they may lack the human touch and emotional understanding necessary for effective patient care. Human oversight ensures that patients receive compassionate care and have a trusted healthcare provider who can address their concerns and provide emotional support.

CHAPTER-4

PROPOSED METHODOLOGY

³ With the help of AI and machine learning techniques, a basic user interface (UI) made with HTML and CSS can be created that enables users to quickly submit their symptoms and obtain a medical diagnosis. Usually, this user interface consists of text fields for users to enter their symptoms, a button to submit the data, and a results section that shows the projected illness. To assess input symptoms and determine the most likely ailment, the system's backend utilizes machine learning algorithms like Random Forest Classifier and XGBoost. ⁴³ The Random Forest method builds multiple decision trees during training and averages their predictions. It categorizes patients as at risk or not for heart disease after dividing the dataset into training and testing groups. ³ The model's performance is evaluated using accuracy, supporting early identification and preventive healthcare practices. For diabetes prediction, a dataset with features like age, BMI, and blood pressure, apart from many others, is used. After preprocessing, ¹³ the data is split into training and testing sets. Random Forest constructs decision trees from various data subsets and combines their predictions. Model evaluation through accuracy helps medical professionals identify diabetes early and treat it effectively. XG Boost begins by loading and analyzing the parkinsons dataset, separating the labels (which indicate the condition of the disease) from the input data (which are features). It partitions ¹⁵ the data into training (85%) and testing (15%) sets and scales the features ³ for consistency. Using the training data, the model is trained to find patterns that may indicate Parkinson's disease. While accuracy is the program's primary statistic, it may also assess precision and recall for a more thorough performance study.

Because the interface is so simple, even people with little technological expertise can readily interact with the system. It is critical to address equity and reduce biases in AI algorithms and data as AI becomes more incorporated into healthcare. Particularly when it comes to underprivileged groups or uncommon diseases, ³⁶ bias in training data or algorithmic decision-making might produce discriminatory results. The objective is to make sure that training data is representative and varied, and that algorithms are routinely examined and tested to find and fix biases. Establishing moral guidelines and international laws is also necessary to

guarantee the ethical and fair application of AI in healthcare. By focusing on data protection, patient permission, openness, and accountability, these policies should guarantee that AI tools serve physicians, benefit patients, and foster healthcare innovation without sacrificing safety or equity.

Dataset used for heart disease prediction: heart.csv

This dataset, which includes demographic data and a variety of health metrics, is associated with heart disease. The number of rows and columns is broken down below, along with a synopsis of each column.

Dataset Dimensions: Number of Rows: 303, Number of Columns: 14.

Columns Overview

Age: The individual's age, expressed in years. This is an ongoing variable that may affect the risk of heart disease.

Sex: The individual's gender (1 = male, 0 = female). Understanding gender-related variations in the prevalence of heart disease is made easier by this categorical variable. Chest pain type, or CP, is a classification of the following types of chest pain:

0: Conventional angina

1. atypical angina

2. Pain that is not angina

3. No symptoms

52

Trestbps: The person's blood pressure at rest (measured in millimeter-Hg). An essential component of evaluating cardiovascular health is this continuous variable.

Chol: The amount of serum cholesterol (in milligrams per deciliter). One important heart disease risk factor is high cholesterol.

The fasting blood sugar level, or Fbs, is as follows: 1 = true if > 120 mg/dL, 0 = false. The existence of diabetes risk is indicated by this binary variable.

Results of a resting electrocardiogram, or "restecg," are divided into the following categories:

0: Normal

First, exhibiting aberrant ST-T waves

2. Exhibiting either clear or likely left ventricular hypertrophy Thalach (attained maximal heart rate): bpm, or the highest heart rate attained during activity. Greater values may be a sign of improved cardiovascular fitness. Exang, or exercise-induced angina, is a metric that indicates whether or not angina occurred during exercise (1 = yes, 0 = no). It is crucial to consider this binary variable when evaluating exercise tolerance.

Oldpeak: Exercise-induced ST depression compared to rest (a cardiac function metric). This

24

School of Computer Science Engineering & Information Science, Presidency University.

constant may be a sign of ischemic heart disease.

Slope: The peak workout ST segment's slope, divided into the following categories:

Upsloping (0), flat (1), and downsloping (2).

Ca, or the number of major vessels colored by fluoroscopy, is the total number of major vessels (0–3) that have undergone fluoroscopy coloring. This variable aids in determining how severe coronary artery disease is.

Thal (thalassemia): classification of thalassemia state, including:

1. Typical
2. Repaired flaw

3: Reversible flaw

The goal variable that indicates whether heart disease is present (1 = presence, 0 = absence) is the target variable. This dataset's main objective is to use the features offered to determine if a patient has heart disease or not. Numerous machine learning methods, including logistic regression, decision trees, Random Forest, and support vector machines, are commonly used to do this.

Dataset used for diabetes disease prediction: diabetes.csv

Predicting diabetes, particularly Type 2 diabetes, is a frequent use for this dataset. It includes a number of characteristics that are important for diabetes diagnosis and treatment, as well as an outcome variable that shows whether diabetes is present or not. There are 768 rows in the dataset, and each row represents a distinct patient. The dataset has eight columns, or characteristics, which are listed below:

Pregnancy: The number of pregnancies the patient has encountered. This characteristic can reveal hormonal shifts that impact glucose metabolism and serve in determining the risk for gestational diabetes.

Glucose: In an oral glucose tolerance test, measure the plasma glucose levels after two hours. Diabetes is mostly indicated by elevated glucose levels. The diagnosis of the illness depends on this measurement.

Blood Pressure: Diastolic blood pressure (mm Hg) is the blood pressure measurement. Diabetes is frequently linked to hypertension, which can raise the probability of complications.

SkinThickness: Skin fold thickness of the triceps (mm). Insulin resistance and body fat may be indirectly indicated by this parameter.

Serum insulin: 2-hour (μ U/ml). Insulin resistance and body fat may be indirectly indicated by this parameter.

30

BMI: Weight in kg/(height in m)² is the body mass index, or BMI. Body mass index is a key health factor for diabetes. Higher BMI values indicates obesity, which is directly correlated to the development of diabetes.

DiabetesPedigreeFunction: A function which utilizes family history to rate a person's risk of acquiring diabetes. This characteristic is crucial for comprehending familial risk factors since it measures the hereditary potential to develop diabetes.

Age: The patient's age in years.

55

With the use of the dataset, a Random Forest Classifier may be used to create a promising diabetes prediction model. A technique for ensemble learning called Random Forest builds 37 many decision trees during training and outputs the individual trees' mode of categorization.

Dataset used for Parkinson's disease prediction: parkinsons.data

The dataset you provide uses a variety of voice characteristics to identify Parkinson's illness. A participant's voice recording is represented by each row, and different metrics pertinent to voice and speech analysis are included in the columns.

Overview Columns for the Dataset Described

Column Name Description Name Recording identifier (phon_R01_S01_1, for example).

MDVP:Fo (Hz)Hertz is the unit of measurement for fundamental frequency (Fo).

The maximum frequency (Fhi) expressed in Hertz is MDVP:Fhi(Hz).

MDVP: Flo (Hz)Hertz for the minimum frequency (Flo).

MDVP: Jitter (%)A metric used to quantify frequency variability is jitter %.

MDVP: Abs jitterAbsolute jitter is a frequency stability metric.

MDVP: RAPAnother jitter metric is Relative Average Perturbation.

MDVP: PPQ₁₄Perturbation of Pitch variability is measured by the quotient.

Jitter: DDPPitch perturbation differences to differences.

MDVP: ShineAmplitude variation in voice, indicating loudness stability.

44

MDVP:Shimmer(dB) Shimmer measured in decibels.

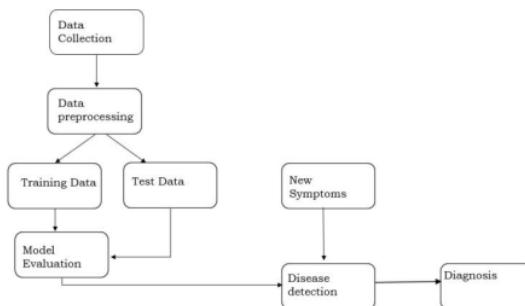
14

Shimmer:APQ₃ Amplitude perturbation

MDVP:APQ (overall amplitude perturbation quotient); SHIMMER:DDA (difference of differences of amplitude perturbation); NHR (noise-to-harmonics ratio), HNR (harmonics-to-noise ratio), status label (presence (1) or absence (0) of Parkinson's disease); RPDE (recurrence period density entropy), a measure of signal complexity; DFA (detrended fluctuation analysis), a measure of signal variability; spread1 (spread of the first set of features); D2 (correlation dimension), a measure of fractal dimension; PPE (pitch period

entropy), which indicates irregularity in pitch.

Architecture:



The initial phase of gathering relevant information is called **data collection**. The collected data is cleaned and prepared for analysis at the **Data Preprocessing** phase. Two preprocessed data sets exist: To train the model, training data is used and Test Data is used to evaluate how well the model performs. In the **Model Evaluation** phase, the model is assessed to determine its predictive accuracy after training. The **New Symptom** branch demonstrates the ability to add more symptoms for analysis to the system. By examining the input symptoms, the model determines potential illnesses in the **Disease detection** phase. The system's final output, "Generate Diagnosis," is a diagnosis based on the user-inputted symptoms and the condition that was found. A simplistic **HTML** and **CSS** **user interface (UI)** that allows users to rapidly input their symptoms and receive a medical diagnosis may be developed with the use of **AI** and **machine learning algorithms**. **Flask** is a web framework designed to be minimalist. To make use of its characteristics, which include `render_template` for controlling **HTML** templates, the `flask` module must be manually imported. At the core of the **Flask** application is the `doctor_app` module, within which the `app` object is assembled. In a nutshell, **Flask** functions as this web application's foundation, handling **HTTP** requests, routing, and dynamically obtaining **HTML** templates. Considering this structure is scalable and adaptable implementing new routes or templates as the program broadens is effortless. **Heart Disease Prediction:** Based on the input dataset, the Python software makes predictions about heart disease using a machine learning model. The steps followed by the program are as follows:

1. **Data Collection:** The application uses ``pd.read_csv('content/heart.csv')`` to read the heart disease dataset from a CSV file. The dataset is explored and understood using the

`'hdata.head()', 'hdata.tail()', 'hdata.shape', 'hdata.info()', 'hdata.isnull().sum()', and
'hdata.describe()'` methods.

2.Data Preprocessing:Using the formulas `'X = hdata.drop(columns='target', axis=1)'` and `'Y = hdata['target']'`, the software divides the dataset into features (X) and target variable (Y).

3.Training Data:`'train_test_split(X, Y, test_size=0.2, random_state=42)'` is used by the software to divide the dataset into training and testing sets. The machine learning model is trained using the training data (`X_train, Y_train`).

4.Test Data: The performance of the trained model is assessed using the testing data (`X_test, Y_test`).

5.Model Evaluation:The program creates a Random Forest Classifier model using `'RandomForestClassifier(n_estimators=100, random_state=42)'`.

- The model is trained on the training data using `'model.fit(X_train, Y_train)'`.
- The model's predictions on the test data are obtained using `'Y_pred = model.predict(X_test)'`.

- The accuracy of the model is calculated using `'accuracy_score(Y_test, Y_pred)'`.

6.New Symptoms:The program creates a new input data point using `'inputData = (75, 0, 2, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1)'`.The input data is converted to a numpy array and reshaped using `'input_array_data = np.asarray(inputData)'` and `'input_data_reshaped = input_array_data.reshape(1, -1)'`.

7.Disease Detection: The trained model is used to predict the class (0 or 1) for the new input data using `'prediction = model.predict(input_data_reshaped)'`

8.Diagnosis: If the prediction is 1, the program prints "The Person has a Heart Disease".If the prediction is 0, the program prints "The Person does not have Heart Disease".

The user interface displays to be an app for doctors called "Doctor App" and is separated into three main sections: Parkinson's Disease, Diabetes, and Heart Disease. The user can enter patient data, including age, cholesterol, resting blood pressure, and maximum heart rate, in the Heart Disease section, and the application will diagnose the patient with heart disease based on the information entered; in the Diabetes section, the user can enter the patient's BMI and glucose level, and the application will diagnose them with diabetes; in the Parkinson's Disease section, the user can enter the patient's MDVP:Fo(Hz) value, and the application will diagnose them with Parkinson's disease.

This Python program, which follows the architecture steps shown in the image, includes data collection, preprocessing, model training and evaluation, and prediction on new data. In

58

summary, it shows how to use a Random Forest Classifier to predict the presence or absence of heart disease based on a given dataset.(Similarly for other diseases such as diabetes and parkinsons).Typically, this user interface includes text areas where users may enter their symptoms, a results section that displays the estimated sickness, and a button to submit the data. The backend of the system uses machine learning methods such as Random Forest Classifier and XGBoost to evaluate input symptoms and identify the most likely condition. During training, the Random Forest technique creates many decision trees and averages their forecasts. After splitting the dataset into training and testing groups, it classifies patients as either at risk or not for heart disease. Accuracy is used to assess the model's performance, which promotes early detection and preventative medical procedures. Among many other variables, a dataset containing blood pressure, age, and BMI is used to predict diabetes. The data is divided into training and testing sets following preprocessing. Using different data subsets, Random Forest builds decision trees and aggregates their forecasts. Accurate model evaluation aids in the early detection and efficient treatment of diabetes by medical experts.

The Parkinson's dataset is loaded and analyzed using XG Boost, which then separates the input data (which are features) from the labels (which show the state of the disease). It scales the features for consistency and separates the data into training (85%) and testing (15%) groups. The model is trained to identify patterns that could point to Parkinson's disease using the training data. The program's main metric is accuracy, but for a more complete performance analysis, it may also evaluate precision and recall.

Even those with less background in technology may easily engage with the system due to its user-friendly interface.

CHAPTER-5

OBJECTIVES

- **Precise Disease Prediction:** The software ought to accurately estimate the sickness or diseases the patient may have depending on the symptoms entered. Make use of complex models that have been trained on big datasets of patient symptoms and disease diagnoses, such as machine learning or expert systems. Both common and unusual diseases should be supported by the system.
- **Simple to Use Interface:** Patients and healthcare professionals should be able to enter symptoms into the software with ease owing to an intuitive interface. Create a straightforward and easy-to-use interface that either lets users enter free-text (symptom description). It should be possible to access the system from a variety of platforms, including desktop and mobile phones.
- **Constant Learning and Improvements:** Keep the software updated with the most recent medical research and changing disease trends, and make sure it learns from fresh data. Provide a way for the machine learning model to be updated and retrained on a regular basis using user input and freshly collected medical data.
- **Security of Data and Privacy:** In accordance with medical data protection laws, make sure patient data is handled with the utmost privacy and security. Protect sensitive health information by using secure data storage techniques.
- **Assistance in Diagnostics for Medical Professionals:** To give medical practitioners a decision support tool or diagnostic recommendations to help them. Make that the software is able to produce predictions along with supplementary data, like diagnostic standards or possible tests required for validation.
- **Adherence to Medical Standards:** Make that the system complies with national and international health regulations and medical diagnostic standards. Update the system periodically to reflect the most recent diagnostic standards, clinical recommendations, and methods for treatment.

By achieving these goals, the system will give patients and medical professionals a dependable and effective tool to aid in the diagnostic procedure. It is crucial to keep in mind that this software should be used as a supplementary aid in medical decision-making rather than as a substitute for a professional diagnosis.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

DESIGNPROCEDURE

- Python libraries: numpy, pandas, matplotlib, seaborn, scaler
- Machine learning algorithms: Random Forest Classifier, XG Boost
- Pandas facilitates data analysis and manipulation, whereas NumPy effectively applies numerical operations on medical data.
- Medical data can be better understood by using data visualization tools like Matplotlib and Seaborn.
- Scalers ensure consistent scales for various properties and preprocess data.
- Medical data can be transformed into predictive models with machine learning techniques like decision trees and support vector machines.³⁹
- These algorithms are able to identify trends, learn from patient data, and accurately predict the diagnosis of diseases.
- Flask is a web-based tool that may be accessed via a browser instead of being a stand-alone desktop program. The application can be used by users from any web-browsing device, making deployment and accessibility simple.

IMPLEMENTATION

Several crucial phases are involved in the construction of the illness prediction program, which combines data collecting, processing, and prediction features. The system begins by creating an intuitive user interface, like a website, where users may enter their symptoms, specific health factors, and optional health information. The backend has programming that maps diseases patterns to indicators and other risk factors. When making predictions, a machine learning model or rule-based algorithm that has been trained on past medical data analyzes the inputs and suggests potential diagnoses along with probability levels. This project makes use of the Random Forest Classifier and XG Boost models. To assess input symptoms and determine the most likely disease, the system's backend employs machine learning techniques like XGBoost and Random Forest Classifier.¹³ The Random Forest method generates a large number of decision trees and averages their predictions.⁴⁵ Patients are classified as either at risk or not for heart disease after the dataset has been divided into training

and testing groups. In order to evaluate the model's effectiveness and encourage early identification and preventative medical measures, accuracy is utilized. To predict diabetes, a dataset comprising age, BMI, and blood pressure is utilized.

Preprocessing is followed by the division of the data into training and testing sets.

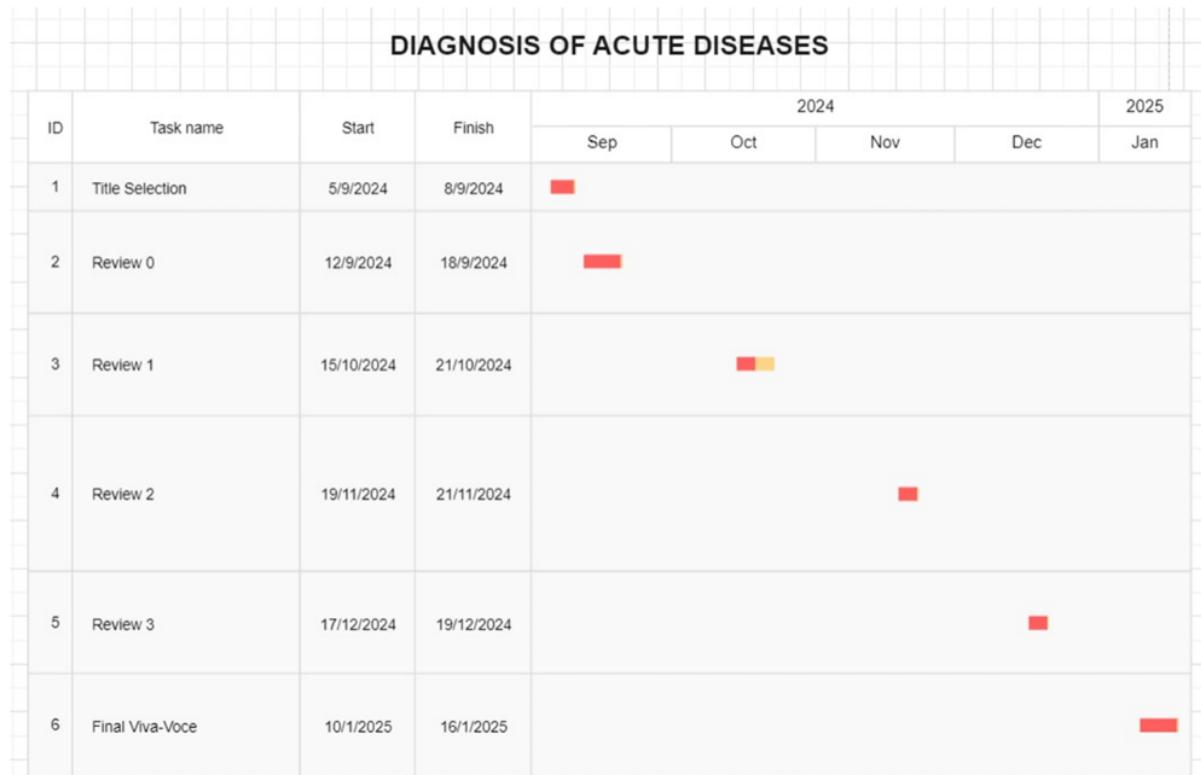
Random Forest constructs decision trees with various data subsets and aggregates their predictions. Precise model assessment facilitates the early identification and effective management of diabetes by health professionals. The Parkinson's dataset is loaded and analyzed using XG Boost, which then separates the input data (which are features) from the labels (which show the state of the disease). It scales the features for consistency and separates the data into training (85%) and testing (15%) groups.

The model is trained to identify patterns that could point to Parkinson's disease using the training data. As a web-based utility that can be used through a browser rather than as a standalone desktop application, Flask makes deployment and accessibility easy because users may use the application from any web-browsing device.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)



CHAPTER-8

OUTCOMES

A straightforward and simple interface created with HTML, CSS, and JavaScript was the project's end outcome. Through hyperlinks on the main page, users can enter conditions for ²⁷ heart disease, diabetes, and Parkinson's disease, and the results are anticipated based upon the symptoms they provide. The software assists patients get a hold of healthcare services, especially in underprivileged places like small towns or rural villages. By making remote access possible, it assures stable availability and promotes early disease detection, enhancing public health overall by lowering the need for critical, late-stage treatments. Cost-effectiveness is one key result, that keeps healthcare affordable for consumers and families. Early recognition eliminates the need for costly treatment options and inpatient stays by minimizing the progression of the disease. Individuals with minimum technical expertise are able to utilize the application. The interface's simplicity prompts users to actively participate in their health, thereby removing any hesitation or concerns regarding health care processes.

CHAPTER-9

RESULTS AND DISCUSSIONS

Several crucial steps are involved in developing a basic illness prediction program, including gathering user input, analyzing the data, and formulating predictions using a machine learning model or predetermined logic. In order to forecast possible acute illnesses, the created program efficiently collects and analyzes user input. Heart Diseases are predicted using the Random Forest classifier with an accuracy of 98.5%, after analyzing a dataset with 1025 entries. The program predicts diabetes with an accuracy of 87.6% using Random Forest, after analyzing a dataset of 768 entries. The program trained an XGBoost model by analyzing a dataset of 195 entries related to Parkinson's illness. The model produced largely accurate predictions, with an accuracy of 96.67%.

The program has several advantages. Through its straightforward interface, individuals without technical expertise can benefit from it. For critical situations, the prediction process' speed and near-real-time findings are crucial. Additionally, the software might be enhanced to incorporate new symptoms or diseases in order to maintain up with advances in medical science because it is scalable.

However, there are some constraints on the program. In difficult or uncommon situations, its accuracy may be hampered because of its dependence on the caliber of the data utilized for training or developing its prediction logic. Furthermore, it is simply a preliminary screening tool and oversimplifies multi-symptom diseases. It cannot replace a professional medical diagnosis. A few modifications are put forward for the future to deal with these kinds of problems. Using machine learning models might improve the precision and flexibility of predictions. Further, ethical concerns including safeguarding user privacy and minimizing biases in training data are crucial to the system's long-term reliability and trustworthiness.

The illness prediction program shows considerable potential as a leading-edge instrument that will enhance healthcare awareness and accessibility. Its fundamental architecture successfully blends sophisticated algorithms with user-friendly features to provide prompt and precise predictions depending on user inputs. Its capacity to facilitate early illness identification has been one of its biggest achievements. Treatment outcomes and implications

can be significantly enhanced through early detection of conditions involving diabetes, heart disease, and Parkinson's. The program serves as a bridge for users who would be reluctant to seek emergency medical treatment because they lack the knowledge, time, or supports to go about so by making predictive insights readily available.

The software's ability to help in initial screening makes it a useful supplement tool in the healthcare ecosystem, even if it cannot replace expert medical guidance. Its contributions to public health research, health awareness, and early detection demonstrate its innovative potential to improve healthcare delivery and results globally. With further development, this instrument may establish itself as a key component in the fusion of technology and medicine, influencing a future in which easily available and customized healthcare is the norm.

CHAPTER-10

CONCLUSION

When creating software that predicts diseases using artificial intelligence (AI) and user-reported symptoms, it is important to understand that this tool should be used as an additional tool to support medical decision-making, not as a substitute for a professional diagnosis.²⁵ The ability to precisely evaluate and handle vast amounts of data from multiple sources, such as patient histories, medical records, and symptom input, has been shown to be possible with AI technologies, especially those that use machine learning. In order to assist in identifying prevalent illnesses and anticipating future health problems, these technologies can swiftly spot patterns and correlations that human doctors could take longer to see. Although AI has the potential to improve medical diagnosis efficiency and accuracy, it is crucial to make sure that these systems are managed by humans. Furthermore, AI models are sometimes criticized for their lack of transparency, which means that they might not always articulate how they arrive at a specific diagnosis.¹⁵ This lack of clarity might make it difficult to trust and hold people accountable, particularly if a diagnosis or recommended course of treatment has unfavorable effects. Maintaining the explainability of AI decisions is essential for boosting system trust and giving medical professionals the knowledge they need to make wise choices. Furthermore, the ethical issues with AI in healthcare extend beyond issues of responsibility and transparency. The possibility of employment displacement is one problem, especially for medical professionals whose jobs could be automated by AI in the diagnostic process. Although AI can help medical personnel, there is a worry that technology may eliminate the need for human labor in some fields, which would change job trends and need retraining for impacted employees. Given the difficulties and moral dilemmas—such as accountability, transparency, and possible job displacement—AI-powered diagnostic software has a lot to offer the medical field. With its ability to analyze user-reported symptoms and suggest possible diseases, it can improve diagnostic speed and accuracy, especially in underprivileged areas with limited access to medical specialists. This can lessen the workload for healthcare professionals, freeing them up to concentrate on more complicated situations, prioritize care, and prioritize patients. Additionally, AI has the potential to continuously enhance its forecasts over time, improving health outcomes. In order to ensure ethical decision-making, the software should always be used under human supervision. However, it can empower patients to seek early medical attention, optimize

resource allocation, and support healthcare professionals, all of which will improve overall healthcare delivery, efficiency, and accessibility.

REFERENCES

- [1] Nafiseh Ghaffar Nia, Erkan Kaplanoglu et al. (2023) ‘Evaluation of artificial intelligence techniques in disease diagnosis and prediction’.
- [2] Mohammed Naseeruddin Taufiq, Bandaru Bhavagna Shreya et al. (2024) ‘Diagnosis Of Acute Diseases In Villages And Smaller Towns Using AI’.
- [3] Madhu H T, Sachin S et al. (2024) ‘Using AI, Diagnosis of Acute Diseases in Villages and Smaller Towns’.
- [4] Feng Shi, Jun Wang et al. (2021), ‘Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19’.
- [5] Panagiota Efthymiou, Athanassios Vozikis et al. (2020) ‘AI and Big Data: A New Paradigm for Decision Making in Healthcare’.
- [6] Ali Husnain, Ayesha Saeed et al. (2024) ‘Harnessing AI for Early Detection of Cardiovascular Diseases: Insights from Predictive Models Using Patient Data’.
- [7] Ugwu Okechukwu P.C., Paul Chima et al. (2024) ‘The Use of AI in Detecting Rare Diseases’.
- [8] Kiswendsida Kisito Kaboré, Omar Sawadogo et al. (2024) ‘Towards a Chatbot for Medical Diagnosis Based on Patient Symptoms’.
- [9] Zeineb Safi, Alaa Abd-Alrazaq et al. (2020) ‘Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review’.
- [10] Sandra Brasil, Carlota Pascoal et al. (2019) ‘Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter?’
- [11] ‘AI in healthcare: Use cases, applications, benefits, solution, AI agents and implementation’.
- [12] Vidhya Rekha Umapathy, Suba Rajinikanth B, (2023) ‘Perspective of Artificial Intelligence in Disease Diagnosis: A Review of Current and Future Endeavours in the Medical Field’.

APPENDIX-A

PSUEDOCODE

1. Import required libraries:

- Import necessary libraries for data handling, preprocessing, modeling, and evaluation.

2. Load Parkinson's dataset:

- Load the dataset from the specified path.

3. Data Preprocessing:

- Drop the 'name' column as it is not a feature.
3
- Separate the dataset into features (X) and the target variable (y).
- Normalize the feature data using MinMaxScaler to scale the values between -1 and 1
7

4. Split the dataset:

- Split the data into training and testing sets (85% training, 15% testing).
60

5. Create and train the XGBoost model:
23

- Create an XGBoost classifier model.
- Fit the model to the training data.

6. Feature Importance Ranking:

- Calculate the feature importance using the trained model.
- Create a ranking of features based on importance.
- Print the sorted feature importance.

7. Select top 4 important features:

- Identify the top 4 most important features based on the ranking.
59
- Create a new dataset using the selected features (including the target variable).

8. Split the dataset again:
6

- Split the new dataset into training and testing sets.

9. Normalize selected features:

- Apply MinMaxScaler to the selected features for normalization.
23

10. Retrain the model:

- Fit the XGBoost model to the training data using the selected features.

11. Evaluate the Model:
2

- Predict the target variable for the test data.
- Calculate the accuracy of the model using accuracy_score.

12. User Input for Prediction:

8

- Define a function to predict Parkinson's disease based on user input.
- The function scales the input data using the same scaler and returns a prediction of "Parkinson's" or "No Parkinson's" based on the model.

13. Get user input:

- Prompt the user to input values for the top 4 important features.

14. Make a Prediction:

- Call the prediction function with the user inputs and print the result.

15. End.

Similarly for other diseases.(Heart disease and Diabetes disease)

APPENDIX-B

SCREENSHOTS

The screenshot shows a web browser window titled "Doctor App" at "localhost:5555". The main header features a circular logo with various medical icons like a stethoscope, a heart, a syringe, and a brain. Below the header, there are three categories: "Heart Disease", "Diabetes", and "Parkinson's". The "Heart Disease" section is active, displaying a form titled "Enter Patient Data" with fields for Age (50), Cholesterol (100), Resting Blood Pressure (80), and Maximum Heart Rate (75). A "Submit" button is present. Below the form, a "Heart Disease Diagnosis" section shows the input values: age: 50 and choli: 100.

The screenshot shows the same web browser window at "localhost:5555". The "Diabetes" section is active, displaying a form titled "Enter Patient Data" with fields for Glucose Level (100) and BMI (80). A "Submit" button is present. Below the form, a "Diabetes Diagnosis" section shows the input values: glucose: 100 and bmi: 80, with a note: "The patient does not have Heart Disease." The "Parkinson's Disease" section is also visible below, showing a similar form and diagnosis message: "The patient is likely to have Diabetes." The "Parkinson's Disease" section has a red background for its input field.

APPENDIX-C

ENCLOSURES

1. International Journal of Innovative Research In Technology (IJRT) has accepted and published the paper.

Reference number is IJIRT171936/Volume 11/Issue 8/







2.The similarity index of the document is 20%.

3.Detailsof mapping the project with the Sustainable Development Goals (SDGs):

26
3. Sustainable Development Goal 3 - Good Health and Well-being:

A major contribution to Goal 3, which is to ensure everyone's health and well-being, is made by this project. The software's straightforward HTML and CSS interface lets users enter disease symptoms, and it uses AI and machine learning algorithms to accurately diagnose the condition. In order to help with early detection and precise treatment planning, this can help medical practitioners by providing possible diagnoses depending on the symptoms entered. The software can improve patient outcomes and the broader healthcare system by increasing the speed and accuracy of diagnosis. This, in turn, helps to improve people's health by giving them fast and accurate health information, which eventually results in better

²⁶
disease and illness management.

Sustainable Development Goal 7 - Affordable and Clean Energy:

This initiative also supports Goal 7, which focuses on clean and economical energy. A large user base from all over the world can access it because it is made to be simple for everyone with an internet connection. The software is compatible with a number of devices, such as laptops, cellphones, and personal computers, and it enables users to receive health diagnostics while they're at home or on the move. The software's internet-based design eliminates the requirement for physical infrastructure and enables users in remote locations or without access to cutting-edge medical facilities to utilize healthcare solutions. By guaranteeing that healthcare solutions are available and cheap for everyone, this promotes the worldwide drive for universal access to high-quality healthcare.

PROJECT_REPORT

ORIGINALITY REPORT



PRIMARY SOURCES

- | | | |
|---|--|------|
| 1 | Submitted to Presidency University
Student Paper | 6% |
| 2 | Submitted to M S Ramaiah University of Applied Sciences
Student Paper | 4% |
| 3 | H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam.
"Computer Science Engineering", CRC Press, 2024
Publication | 1% |
| 4 | Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, Dinggang Shen. "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19", IEEE Reviews in Biomedical Engineering, 2020
Publication | <1 % |
| 5 | www.onlinescientificresearch.com
Internet Source | <1 % |

6	Internet Source	<1 %
7	www.coursehero.com Internet Source	<1 %
8	"Advances in Artificial Intelligence and Machine Learning in Big Data Processing", Springer Science and Business Media LLC, 2025 Publication	<1 %
9	www.ijert.org Internet Source	<1 %
10	Submitted to Pathfinder Enterprises Student Paper	<1 %
11	algovista.org Internet Source	<1 %
12	www.frontiersin.org Internet Source	<1 %
13	"Intelligent Computing Systems and Applications", Springer Science and Business Media LLC, 2024 Publication	<1 %
14	Submitted to Coventry University Student Paper	<1 %
15	T. Mariprasath, Kumar Reddy Cheepati, Marco Rivera. "Practical Guide to Machine	<1 %

Learning, NLP, and Generative AI: Libraries, Algorithms, and Applications", River Publishers, 2024

Publication

- | | | |
|----|--|------|
| 16 | machinelearningtutorials.org | <1 % |
| 17 | Submitted to University of East London | <1 % |
| 18 | ijrpr.com | <1 % |
| 19 | Submitted to Australian Catholic University | <1 % |
| 20 | fastercapital.com | <1 % |
| 21 | Submitted to Arab Open University | <1 % |
| 22 | Bittu, Pavani. "Prediction of the Need for Mental Health Therapy Among COVID-19 Patients", Southern University and Agricultural and Mechanical College, 2024 | <1 % |
| 23 | dokumen.pub | <1 % |
| 24 | ijsred.com | <1 % |

25	www.leewayhertz.com Internet Source	<1 %
26	www.ncri.in Internet Source	<1 %
27	journal.ijresm.com Internet Source	<1 %
28	peerj.com Internet Source	<1 %
29	www.ijraset.com Internet Source	<1 %
30	www.drhealthadvices.com Internet Source	<1 %
31	Submitted to University of Birmingham Student Paper	<1 %
32	Sudhanshu Kumar, Himanshu Yadav, Mandeep Kaur, Mayank Kumar Goyal. "Chapter 13 Machine Learning-Based Cardiovascular Disease Prediction", Springer Science and Business Media LLC, 2024 Publication	<1 %
33	jbid.sciforce.org Internet Source	<1 %
34	www.booksfree.org Internet Source	<1 %

35	5dok.net Internet Source	<1 %
36	Baseer UI Haq, Mohammad Faisal, Mahmood UI Haq. "chapter 4 Ethical Considerations in Remote Sensing and Cognitive Modeling", IGI Global, 2024 Publication	<1 %
37	C. Ashwini, S. Rubin Bose, M. S. Deepika Padmavathy, Calvin Raj, J. Chalwin Ajay. "chapter 17 Predicting Child Mortality With Diverse Regression Algorithms Using a Machine Learning Approach", IGI Global, 2024 Publication	<1 %
38	Natasa Kleanthous, Abir Hussain. "Machine Learning in Farm Animal Behavior using Python", CRC Press, 2025 Publication	<1 %
39	Xinyuan Song, HSIEH,WEI-CHE, Ziqian Bi, Chuanqi Jiang, Junyu Liu, Benji Peng, Sen Zhang, Xuanhe Pan, Jiawei Xu, Jinlang Wang. "A Comprehensive Guide to Explainable AI: From Classical Models to LLMs", Open Science Framework, 2024 Publication	<1 %
40	ejaet.com Internet Source	<1 %

41	Internet Source	<1 %
42	www.ijsr.com Internet Source	<1 %
43	Arvind Dagur, Dhirendra Kumar Shukla, Nazarov Fayzullo Makhmadiyarovich, Akhatov Akmal Rustamovich, Jabborov Jamol Sindorovich. "Artificial Intelligence and Information Technologies", CRC Press, 2024 Publication	<1 %
44	Submitted to Liverpool John Moores University Student Paper	<1 %
45	Pethuru Raj, N. Gayathri, G. Jasper Willsie Kathrine. "Artificial Intelligence for Precision Agriculture", CRC Press, 2024 Publication	<1 %
46	Suhel Ahmad Khan, Mohammad Faisal, Nawaf Alharbe, Rajeev Kumar, Raees Ahmad Khan. "Computational Intelligence Applications in Cyber Security", Routledge, 2024 Publication	<1 %
47	qubixity.net Internet Source	<1 %
48	www.chromatographyonline.com Internet Source	<1 %

- 49 www.geeksforgeeks.org <1 %
Internet Source
-
- 50 10www.easychair.org <1 %
Internet Source
-
- 51 A. Sivasangari, G.V. Sai Kanth, Ch. Sai Deekshith, G. Manohar, B. Arun Kumar, K. Vasavi. "A Review on Infectious and Cardiovascular Disease Classification Using Deep Learning", 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023 <1 %
Publication
-
- 52 Arslan Naseer, Muhammad Muheet Khan, Fahim Arif, Waseem Iqbal, Awais Ahmad, Ijaz Ahmad. " An improved hybrid model for cardiovascular disease detection using machine learning in ", Expert Systems, 2023 <1 %
Publication
-
- 53 Ikome, Otto M.. "Allostatic Load as an Early Diagnostic Tool and Predictor of Cardiovascular Disease: Using an Integrated Gompertz Logistic Random Forest Model.", Jackson State University, 2024 <1 %
Publication
-
- 54 Poonam Tanwar, Tapas Kumar, K. Kalaiselvi, Haider Raza, Seema Rawat. "Predictive Data <1 %

Modelling for Biomedical Data and", River
Publishers, 2024

Publication

55	jlacollectivo.com Internet Source	<1 %
56	magnascientiapub.com Internet Source	<1 %
57	stackoverflow.com Internet Source	<1 %
58	www.racgp.org.au Internet Source	<1 %
59	Joice. C Sheeba, M. Selvi. "Pedagogical Revelations and Emerging Trends", CRC Press, 2025 Publication	<1 %
60	Di Wu. "Data Mining with Python - Theory, Application, and Case Studies", CRC Press, 2024 Publication	<1 %
61	Thompson Stephan. "Artificial Intelligence in Medicine", CRC Press, 2024 Publication	<1 %
62	Tsakonas, A.. "An evolutionary system for neural logic networks using genetic programming and indirect encoding", Journal of Applied Logic, 200409 Publication	<1 %

Exclude quotes Off
Exclude bibliography On

Exclude matches Off