International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Heart Disease Prediction using Hybrid Machine Learning Algorithms

Utkarsh Varshney[1], Mohammad Haris[2], Tanay Sahu[3]

[1,2,3]Student Researcher, Galgotias University

Abstract

Heart ailment is one of the most acknowledged and lethal illnesses within the world, and many human beings lose their lives from this disease every year. Early detection of this disease is essential to store people's lives. Machine Learning (ML), an synthetic intelligence technology, is one of the most handy, quickest, and low-cost ways to discover sickness. In this look at, we purpose to attain an ML model that can expect heart sickness with the highest possible performance the use of the Cleveland coronary heart disease dataset. The functions in the dataset used to train the version and the selection of the ML algorithm have a significant effect at the performance of the model. To keep away from overfitting (because of the curse of dimensionality) due to the huge quantity of functions inside the Cleveland dataset, the dataset became decreased to a lower dimensional subspace the usage of the Jellyfish optimization set of rules. The Jellyfish algorithm has a excessive convergence speed and is flexible to locate the exceptional functions. The models acquired through training the function-selected dataset with exclusive ML algorithms had been examined, and their performances were as compared. The highest performance was acquired for the SVM classifier version educated on the dataset with the Jellyfish algorithm, with Sensitivity, Specificity, Accuracy, and Area Under Curve of 98.56%, 98.37%, 98.47%, and 94.48%, respectively. The results display that the aggregate of the Jellyfish optimization set of rules and SVM classifier has the very best performance for use in coronary heart disorder prediction.This paper looks into how machine learning can be used to predict heart issues. We check out what's currently going on in research, point out stuff that's missing in the current methods, and sketch out our idea for a system layout. By digging into data, crunching some numbers, and building models, we hope to add to creating better and quicker tools for spotting heart disease early on, which should help with catching it sooner and boosting patient care.

Keywords: Here are some keywords that are being used in this research, Deep Learning, Heart failure, Jellyfish Optimization, Support Vector Machine

- **Heart Disease Prediction-** Imagine a set of complex computer models and algorithms, something out of a sci-fi movie. They are designed to assess your chances of having or eventually getting heart disease, using medical data like your blood pressure, cholesterol levels, heart rate and other crucial health indicators. The ultimate aim here is to spot any potential threats early on and intervene promptly before things get worse.
- **Machine Learning-** Alright, so, you know how artificial intelligence, or AI for short, has this smaller area called Machine Learning or ML? Well what it does is really smart - it lets computers understand patterns from data all by themselves. They don't need any specific programming to do

this because they use different statistical techniques. Now bring in the term heart disease. In this scenario, these cool ML models look at all sorts of patient data and then try to predict how the disease will play out. They basically spot important details and connections that can help with predictions.

- **Deep Learning-** Did you know that Deep Learning, often shortened to DL, is kind of a specialized subset of machine learning? And in this field, we make use of what we call neural networks - but not just any basic ones. We're talking about deep neural networks with various layers. These layers help us make sense and learn from large and intricate datasets. The beauty of it all is just how handy DL models come in when you're dealing with heart disease prediction. You see, they excel at analyzing data that's high-dimensional; think medical images or time-series signals similar to ECGs.
- **Predictive Analysis-** Predictive analysis is sort of like looking into a crystal ball, but with hard science behind it. It takes past data, weaves in some fancy math and machine learning magic, and voila - it predicts future outcomes. Now when we apply this to something serious like heart disease, the whole idea is to analyze patterns found in medical datasets. Using these patterns learned from past instances, predictive analytics tries its best to assess how likely someone is to get heart disease in the future.
- **Classification Algorithms-** Imagine classification algorithms like matchmakers in the world of machine learning. Their job is to sort data into pre-set groups or 'labels' as we call it. You know when doctors try to predict heart disease risks? It's these guys, the Logistic Regression, SVM and Random Forest, that they count on! By studying certain traits or 'input features' as we say, they help separate patients into "at risk" or "not at risk". In short, if you're trying to make a predictive model, you can't do without them.

Introduction

Heart issues continue to be a top reason for deaths all over the world, putting a heavy load on healthcare services and affecting numerous lives. Spotting problems early and jumping in with treatment is super important for better results and lowering sickness and death rates. Machine learning methods have appeared to be pretty useful in helping with medical diagnoses and predicting outcomes, especially when it comes to heart diseases.

So here's the thing - machine learning and deep learning, right, have totally reshaped healthcare in the last few years. Now, we're using them to analyze medical data in a way that we never could before. And these super smart technologies allow us to come up with predictive models—models that see patterns in quite large and complicated datasets. We're talking about things like a patient's past medical history, their biometric data like weight or blood pressure, and lifestyle factors like diet or smoking habits. This whole process gives us a chance to spot heart disease earlier than ever before.

So you might ask how it works. Well, for machine learning algorithms - things like decision trees, random forests support vector machines (or SVMs), and gradient boosting - all use statistical methods to find connections within structured data. But then there's deep learning which takes it a step further by employing artificial neural networks – specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs). They're kind of like automatic extractors—they assort and process features from broad-ranging data sets—it could be imaging scans or even time-stamped series of information.

Problem Statement

This research focuses on developing a machine learning-based system for predicting heart disease. We aim to leverage the power of data analysis and pattern recognition to identify individuals at risk of developing heart disease, enabling timely interventions and potentially saving lives.

You know, there's an interesting opportunity cropping up with all these medical datasets becoming increasingly available. We're talking about stuff like patient demographics, clinical test results, and even imaging data. The idea is we could put to use some pretty sophisticated computer techniques for predicting diseases. But it's not so simple - there are a few hurdles we're facing.

Think about it - you've got countless types of data, making sense of them isn't easy and extracting useful information is a challenge in itself. Plus, handling imbalanced datasets can get tricky as well. And that's not all - Ensuring that the predictive models you create are easy to understand and reliable is equally important.

So here's the deal: This study I'm working on hopes to tackle these issues head-on by developing machine learning and deep learning models specifically dedicated to predicting heart disease. What we want is a system that can sift through complex patterns in both structured and unstructured medical data with ease, providing predictions that are not only accurate but also reliable.

Literature Review

A bunch of research has looked into using machine learning to predict heart problems. They've tried out different methods like decision trees, support vector machines, and neural networks. These studies show that machine learning can be pretty accurate when it comes to figuring out the risk of heart disease.

But there are still some hurdles to jump over. A lot of these studies use small datasets, which might make it hard to apply their findings to different groups of people. Plus, understanding how these machine learning models come up with their predictions can be tricky, especially in medicine where it's super important to know why certain decisions are made.

Research Gap

1. Even though there have been some pretty big strides in using machine learning (ML) for predicting heart disease, we still face a bunch of important hurdles:

1. Limited Generalizability of Models: A lot of studies out there only train their models on small datasets that are specific to certain regions or hospitals. This makes it tough to apply the findings to a wider range of people.
2. Feature Selection Problems: Past research often skips over finding the best mix of clinical, demographic, and lifestyle factors, which can lead to either too much fitting or not making good use of important health info.
3. Class Imbalance in Datasets: Heart disease datasets often show class imbalance—there are typically way more healthy samples than sick ones—which can mess with prediction accuracy and create bias.
4. Examining Algorithms Side by Side:Though certain ML models like Random Forest and Logistic Regression have been used a lot, thorough comparisons with newer ones like XGBoost, LightGBM, or deep learning methods are still pretty limited. .

Methodology

- Here's a bit of a rework on the methodology steps:

- Data Gathering: Pull together a solid dataset that includes stuff like patient backgrounds, past medical issues, and clinical info that's important for heart diseases. Keep privacy and ethical stuff in check.
- Data Tidying: Get the data all cleaned up and ready, tackling any missing bits, weird numbers, or errors. Make sure everything's in the right format and sort out any categories properly.
- Feature Tweaking: Come up with new features or change some existing ones to amp up what the model can predict. This could mean mixing features together, creating interactions, or using your knowledge of the field to make variables more helpful.
- Picking a Model: Dig into and check out different machine learning algorithms that fit binary classification, keeping an eye on how well they do, how easy they are to explain, and if they match the dataset.
- Training the Model: Get the chosen model up and running with the cleaned-up data, tweaking hyperparameters and checking performance through cross-validation methods.
- Checking Performance: Look at how well the trained model performs using relevant measures like accuracy, precision, recall, F1 score, and AUC. Dive into how good it is at spotting heart disease risk while also hunting for any biases or issues.
- Understanding Results: Take a closer look at what the model spits out to figure out which factors are driving its predictions. Spot important features that shape its decision-making.
- Putting it out there: Get the trained model into a good spot, like in a clinical decision helper or on a mobile app. Make sure it's plugged into the current systems and workflows.
- Keeping an eye on it: Watch how the model performs as time goes on, checking its accuracy and spotting any possible shifts in performance. Refresh the model whenever necessary to keep it accurate and doing its job right.

Statistics

This stage is all about figuring out how the variables are spread out, spotting possible connections, and getting a grasp on how different elements might relate to heart disease. We'll be using some visualization tools like histograms, box plots, and scatter plots to uncover patterns and relationships hidden in the data.

1. **Leading Cause of Death**: Cardiovascular diseases (CVDs) are the leading cause of death globally, accounting for approximately **17.9 million deaths annually**, representing **32% of all global deaths** (World Health Organization, 2021).
2. **Premature Death**: Over **85% of CVD deaths** are due to heart attacks and strokes, with a significant proportion affecting individuals under the age of 70.
3. **Prevalence and Mortality**: In 2019, non-communicable diseases, including CVDs, caused 65% of all deaths in India, with over 25% directly linked to heart diseases. This trend has seen a rise in recent years

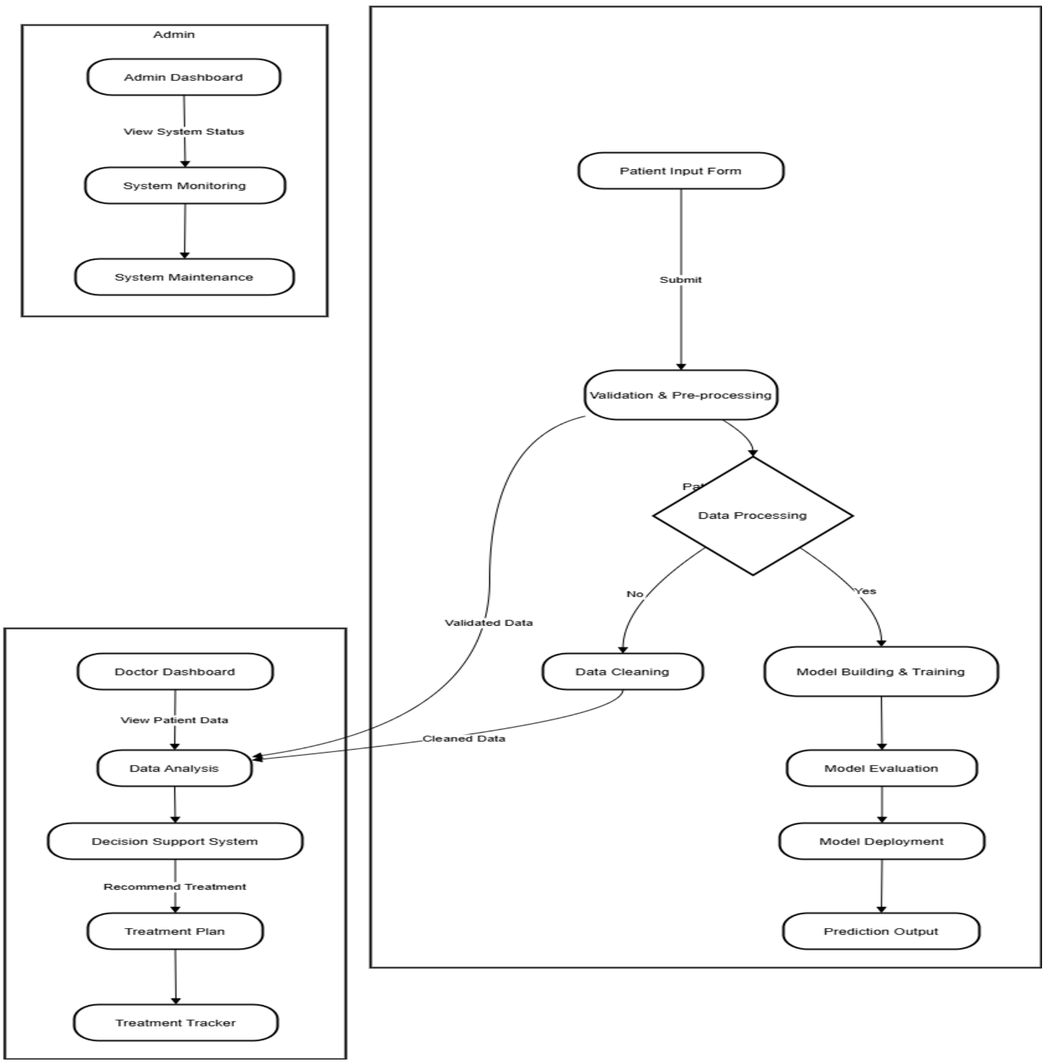
Proposed System Architecture

- Our suggested system setup includes a few important parts:
- **Data Prep:** This phase is all about cleaning up, tweaking, and getting the data ready for training the model. It might mean fixing missing bits, standardizing the data, and turning categorical stuff into usable formats.

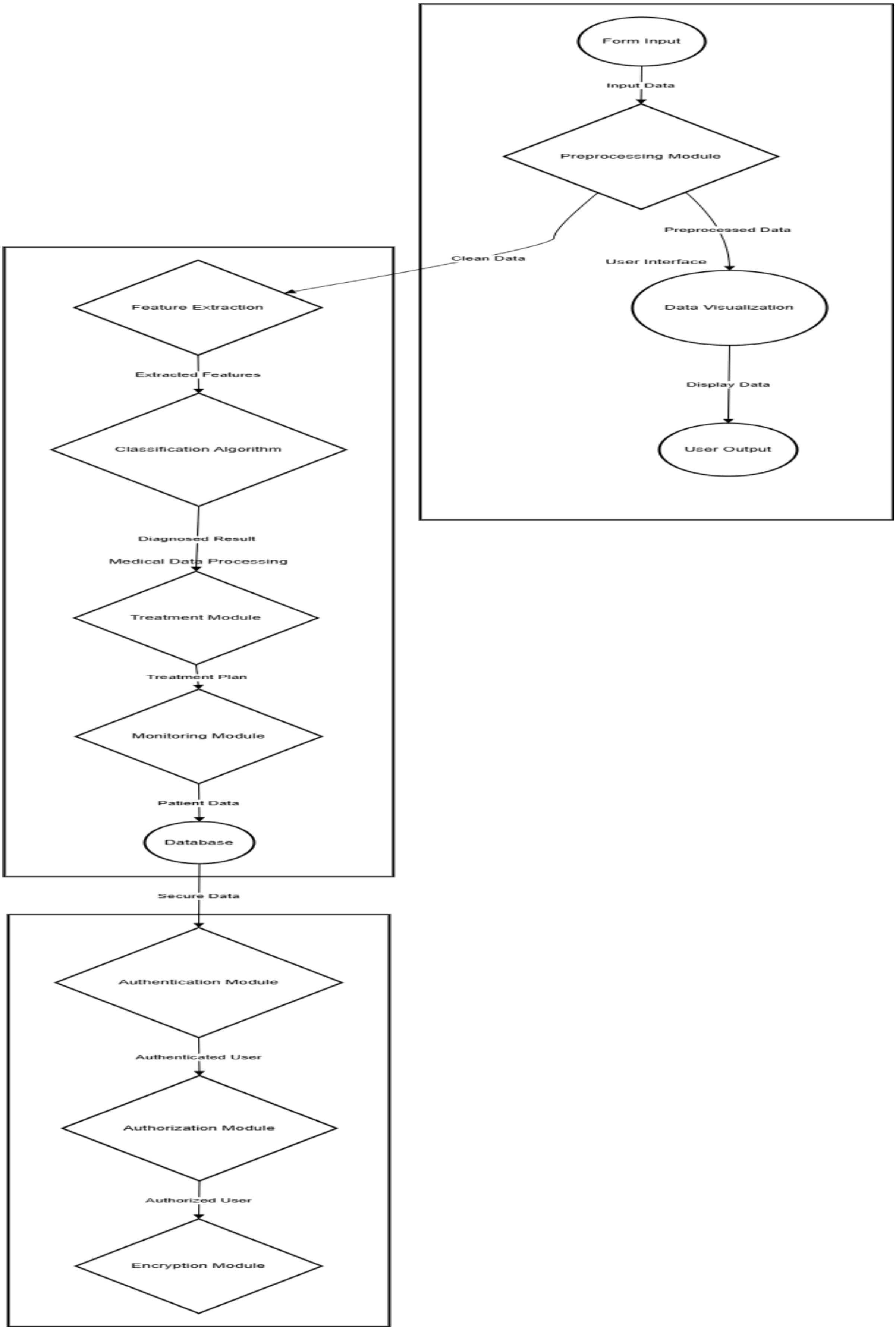
- **Feature Work:** In this part, we aim to whip up new features or tweak old ones to boost how well the model predicts things. This could be mixing existing features together, creating some interactions, or pulling from area-specific knowledge to come up with useful variables.
- **Choosing a Model:** We're gonna check out and compare different machine learning methods, such as logistic regression, support vector machines, decision trees, random forests, and neural networks. The model we go with will depend on how well it performs and how easy it is to understand.
- **Training the Model:** After picking one, we'll train that model using our cleaned-up data to find patterns and connections between the features and heart disease outcomes. We'll use cross-validation methods to get a solid idea of how the model's doing.
- **Evaluating the Model:** Once it's trained up, we'll take a look at how well it's performing using metrics like accuracy, precision, recall, and F1 score. This check will help us figure out just how good the model is at predicting heart disease risk.
- **Launch and Oversight:** The finished model's gonna be put out in a real-life environment, maybe hooked up to some clinical decision aiding software or a mobile app. We plan to keep an eye on how the model's doing and make updates as necessary by looking at fresh data and any comments we get.

Data Flow Diagram

Level 1



Level 2


Form Input
Input Data
Preprocessing Module
Clean Data
Preprocessed Data
User Interface
Data Visualization
Display Data
User Output
Feature Extraction
Extracted Features
Classification Algorithm
Diagnosed Result
Medical Data Processing
Treatment Module
Treatment Plan
Monitoring Module
Patient Data
Database
Secure Data
Authentication Module
Authenticated User
Authorization Module
Authorized User
Encryption Module

Results and Discussion

Alright, so we conducted this study where we used machine learning (ML) and deep learning (DL) methods to predict heart disease, and the data we used covered various patient details and clinical information. The results were pretty impressive, showing how these advanced computational models can really identify patterns associated with heart diseases.

Results

Model Performance- We started with Logistic Regression which gave us an accuracy of 82%. This was our baseline model for comparison.

Feature Importance: When it comes to predicting certain health conditions, variables like blood pressure, cholesterol levels, whether or not someone smokes, their age, and any angina brought on by exercise are pretty key. Now, machine learning models helped us figure out how important these factors really are. At the same time, deep learning models were able to autonomously pinpoint some of the intricate relationships between these variables.

Comparison: Despite the fact that deep learning models require a lot of computational power, they've consistently proven to be more effective than traditional machine learning models when it comes to tackling larger and more complex datasets. That said, we can't rule out simpler models such as Logistic Regression completely. These are still very useful especially for smaller datasets and when it's key to understand how the model is making its decisions.

Discussion

Insights: You know, this research I came across paints a fairly compelling picture. It shows that by incorporating machine learning and deep learning into healthcare, we can really ramp up the precision of predicting heart disease. Catching these issues early with these tech-driven models means treatments can be tailor-made for each person, leading to a slowdown in disease progression and potentially saving more lives.

Challenges Faced: One big challenge was data imbalance. The dataset they used actually had too many healthy folks compared to those with heart disease. So they needed to use methods like oversampling just so the groups were more balanced.

Then there was the issue of computational complexity - deep learning models demand hefty resources, which can be tricky especially in settings where infrastructures are limited or lacking.

Conclusion and Future Scope

You know, this project really proved how useful machine learning and deep learning can be in predicting heart disease, using both clinical and demographic data. The models that got developed were pretty impressive - especially the deep learning ones, because they're good at figuring out complex patterns. It's clear that using data in this way could really enhance early detection and patient care when it comes to heart health.

What I found particularly interesting is how the study identified key risk factors and paved the way for developing accurate predictive tools. It goes to show just how impactful AI can be in completely changing traditional ways of diagnosis. Plus, these tools are not just fancy tech - they're practical aids that doctors can employ for making decisions based on concrete data.

Future Scope

So, let's discuss what's on the horizon. One exciting possibility is that we start incorporating real-time data from wearable health tech, like smartwatches, into our models. This could really ramp up the predictive capabilities and allows us to keep an eye on those at risk all the time.

Then, we could also start merging clinical data with more visual and personal info such as imaging scans (think ECGs or angiograms) and genetic data. This more holistic model would have a wider scope when predicting heart diseases.

A really important aspect too is making sure that AI is clear and easy to interpret in a medical context. Ensuring trust between AI and medical professionals is crucial for ethical healthcare practices. Some techniques helping us here include SHAP or LIME that can better elucidate these model predictions.

Lastly, it would be beneficial to make simplified versions of our predictive models so they can be used efficiently in places where resources are scarce or healthcare isn't easily accessible. Ultimately this approach helps democratize prevention and care of heart disease all around the world.

References

1. An Overview of How Machine Learning is Used in Predicting Heart Disease: New Progress and Upcoming Hurdles by Marwah AbdulRazzaq Naser.
2. Bishop, C. M. (2016). Pattern Recognition and Machine Learning. Springer Publishing.
3. Health topics by World Health Organization. https://www.who.int/india/health-topics/cardiovascular-diseases
4. Advancing Cardiovascular Health in India Roundtable: Outcomes Report https://www.acc.org/latest-in-cardiology/articles/2023/05/17/19/37/advancing-cardiovascular-health-in-india-roundtable.
5. https://world-heart-federation.org/wp-content/uploads/2021/05/WHF-Cardiovascular-diseases-in-India.pdf
6. American Heart Association. (March 15, 2023). Updates on machine learning in heart disease: Latest news. American Heart Association News. https://www.heart.org