CARDIOVASCULAR DISEASE DETECTION USING OPTIMAL FEATURE SELECTION

K. NEHAS REDDY - 21951A04B6

N. MANISH - 21951A0492

B.SNEHA – 21951A04K4

CARDIOVASCULAR DISEASE DETECTION USING OPTIMAL FEATURE SELECTION

A Project Report
submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY IN

ELECTRONICS AND COMMUNICATION ENGINEERING

by

K. NEHAS REDDY (21951A04B6)

N. MANISHKUMAR (21951A0492)

B. SNEHA (21951A04K4)

Under the guidance of

Dr. G MARY SWARNALATHA

Assistant Professor

Department of Electronics and Communication Engineering

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)

Dundigal, Hyderabad – 500 043, Telangana

May 2025

DECLARATION

We certify that

a) The work contained in this report is original and has been done by us under the guidance of my supervisor(s).

b) The work has not been submitted to any other Institute for any degree or diploma.

c) We have followed the guidelines provided by the Institute in preparing the report.

d) We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e) Whenever we have used materials (data, theoretical analysis, figures, and text) from other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references. Further, we have taken permission from the copyright owners of the sources, wherever necessary.

Place: HYDERABAD

Date:

Signature of the Student

21951A04B6

21951A04K4

21951A0492

CERTIFICATE

This is to certify that the project report entitled **CARDIOVASCULAR DISEASE DETECTION USING OPTIMAL FEATURE SELECTION** submitted by team **K. NEHAS REDDY (21951A04B6)** , **N. MANISH KUMAR (21951A0492)** and ,**B. SNEHA (21951A04K4)** to the Institute of Aeronautical Engineering, Hyderabad in partial fulfilment of the requirements for the award of the Degree **Bachelor of Technology in Electronics and Communication Engineering is a bonafide record of** work carried out by them under the guidance and supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor	**Head of the Department**
Dr. G. Mary Swarnalatha	**Dr. P Munaswamy**
Assistant Professor	**Professor**

Date:

APPROVAL SHEET

This project report done **CARDIOVASCULAR DISEASE DETECTION USING OPTIMAL FEATURE SELECTION** by **K. NEHAS REDDY (21951A04B6), N.MANISH KUMAR (21951A0492), B. SNEHA (21951A04K4)** is approved for the award of the Degree **Bachelor of Technology** in **ELECTRONICS AND COMMUNICATION ENGINEERING**

Examiner(s)

Supervisor
Dr. G. Mary Swarnalatha
Assistant Professor

Principal
Dr. L V Narasimha Prasad

Date:

Place: HYDERABAD

ACKNOWLEDGEMENT

We wish to take this opportunity to express a deep gratitude to all those who helped, encouraged, motivated and have extended their cooperation in various ways during my project work. It is our pleasure to acknowledge the help of all those individuals and our family support responsible for foreseeing the successful completion of my project.

We would like to thank my project guide **Dr. G MARY SWARNALATHA, Assistant Professor of Electronics and Communication Engineering** and express my gratitude to **Dr. P MUNASWAMY, Head of the Department** with great administration and respect for their valuable advice and help throughout the development of this project by providing with required information without whose guidance, cooperation and encouragement, this project couldn't have been materialized.

We express our sincere gratitude to **Dr. L. V. Narasimha Prasad, Professor** and **Principal** who has been a great source of information for our work.

We thank our college management and respected **Sri M. Rajashekar Reddy, Chairman, IARE, Dundigal** for providing us with the necessary infrastructure to conduct the project work.

We take this opportunity to express our deepest gratitude to one and all who directly or indirectly helped me in bringing this effort to present form.

ABSTRACT

Cardiovascular disease (CVD) continues to be a cause of death underscoring the pressing need, for effective early detection methods. This study presents a machine learning driven framework for CVD detection focusing on enhancing feature selection from electrocardiogram (ECG) signals. The new system utilizes a range of feature selection techniques, including Fast Correlation Based Filter (FCBF) Minimum Redundancy Maximum Relevance (mRMR) Relief and Particle Swarm Optimization (PSO). These combined techniques are aimed at identifying features for precise classification thereby improving the efficiency of the diagnostic process. The key strength of this framework lies in its feature selection approach. FCBF is employed to eliminate redundant features from the dataset. MRMR further enhances this process by selecting features with relevance to the target variable while minimizing redundancy among them. Relief, a method for weighting features evaluates feature importance based on their ability to differentiate values, between related instances. Finally, PSO optimization fine tunes the feature set by mimicking social behavior patterns like bird flocking to determine the subset of features. The architecture uses Extra Trees (Trees) and Random Forest classifiers to categorize the optimized features. These ensemble learning methods are recognized for their reliability and precision, in managing datasets. The Extra Trees classifier, with its randomized selection of splits and averaging of outcomes is beneficial, for decreasing variability and preventing overfitting. Random Forest, which comprises decision trees, enhances prediction accuracy by combining the results of multiple trees and mitigating the risk of overfitting.The combination of these classifiers within the proposed system achieves remarkable accuracy rates of 100%, demonstrating its efficacy in early CVD detection. Such high accuracy is indicative of the system's potential to significantly improve diagnostic processes in healthcare settings. A comprehensive comparative analysis with state-of-the-art methods was conducted to validate the effectiveness of the proposed approach. This analysis involved diverse datasets to ensure that the system is versatile and generalizable across different types of ECG data. The results consistently showed that the proposed architecture outperforms existing methods, confirming its superiority in feature selection and classification accuracy.

Keywords: Cardiovascular Disease(CVD), Decision trees, random forest**.**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

Cardiovascular disease (CVD) remains one of the leading causes of morbidity and mortality worldwide, accounting for a significant proportion of deaths annually. Despite advancements in medical science, early detection and accurate diagnosis of cardiovascular conditions remain crucial for effective management and treatment. One promising approach to improving the detection and prognosis of cardiovascular diseases is the application of Machine Learning (ML) techniques, particularly through the use of optimal feature selection methods. These methods enhance the predictive power and efficiency of diagnostic models, providing a significant edge in clinical settings.

Feature selection plays a pivotal role in the realm of machine learning and data analytics, particularly in the medical field where datasets are often vast and complex. In essence, feature selection involves identifying and selecting the most relevant and informative variables (features) from a dataset, which are then used to build predictive models. This process is crucial for several reasons: it helps in reducing the dimensionality of the data, minimizes overfitting, enhances model interpretability, and ultimately improves the overall performance of the prediction models. In the context of cardiovascular disease detection, optimal feature selection can lead to more accurate and reliable identification of individuals at risk, thereby facilitating timely interventions.

Cardiovascular diseases encompass a wide range of conditions affecting the heart and blood vessels, including coronary artery disease, heart failure, arrhythmias, and more. These conditions are influenced by a multitude of factors, both genetic and environmental, making the prediction and diagnosis of CVDs inherently complex. Traditional diagnostic methods, while effective, often rely heavily on invasive procedures and can sometimes fall short in predicting the onset of diseases in asymptomatic individuals. This is where machine learning and optimal feature selection come into play, offering a non-invasive, data-driven approach to identify potential cardiovascular issues before they become critical.

The application of machine learning in CVD detection involves the utilization of various algorithms to analyze and interpret medical data, which can include patient demographics,

1.2 OBJECTIVES

- **Statistical Property of Each Feature of Small Data**

  - Examine the statistical properties, such as mean, median, standard deviation, and range, for each feature in the small dataset to understand their individual distributions and central tendencies.

- **Distribution of Numerical Features**

  - Analyze the distribution of numerical features to identify patterns, skewness, and potential outliers. This can be visualized through histograms, box plots, and density plots.

- **Accuracy of All Models on Small Dataset**

  - Evaluate the performance of various machine learning models on the small dataset. This includes assessing metrics such as accuracy, precision, recall, and F1-score for each model.

- **ROC Curves for MrMr, FCBF, Lasso, Relief, and ANOVA**

  - Generate and analyze Receiver Operating Characteristic (ROC) curves for models employing different feature selection techniques such as Minimum Redundancy Maximum Relevance (MrMr), Fast Correlation-Based Filter (FCBF), Lasso, Relief, and ANOVA. Compare the Area Under the Curve (AUC) to determine the effectiveness of each technique.

- **Overall Results of All Classifiers with Confusion Matrix**

  - Compile and present the overall results of all classifiers using confusion matrices. This will help in understanding the true positives, false positives, true negatives, and false negatives for each model.

- **Accuracy of Each Model on Each Selection Technique**

    - Compare the accuracy of each model when different feature selection techniques are applied. This involves analyzing how each method impacts the model's predictive performance.

- **Pearson Correlation Between All the Features**

    - Calculate the Pearson correlation coefficients between all pairs of features to assess the degree of linear correlation. This can help in identifying redundant features and understanding feature interdependencies.

1.3 FEASIBILITY

The feasibility of utilizing optimal feature selection for cardiovascular disease (CVD) detection is grounded in the convergence of several critical factors, including advancements in data collection, the proliferation of machine learning algorithms, and the increasing availability of computational resources. These elements collectively create a conducive environment for implementing sophisticated analytical techniques in clinical settings, potentially transforming the landscape of CVD diagnostics and personalized medicine.

Firstly, the vast and growing availability of health data plays a pivotal role in the feasibility of this approach. Electronic health records (EHRs), wearable health devices, and other sources generate an immense amount of data that can be leveraged for predictive modeling. EHRs provide comprehensive patient information, including demographics, medical history, laboratory results, and imaging studies. Wearable devices offer continuous monitoring of physiological parameters such as heart rate, blood pressure, and activity levels. This wealth of data is a valuable resource for developing robust machine learning models, provided that it is appropriately curated and preprocessed.

The rapid advancement of machine learning techniques further enhances the feasibility of optimal feature selection for CVD detection. Machine learning algorithms have demonstrated remarkable success in various domains, including image recognition, natural language processing, and predictive analytics. In the context of CVD detection, algorithms such as logistic regression, decision trees, support vector machines, and neural networks can be employed to build predictive models. These models can analyze complex interactions among

Feature selection methods are an integral component of this analytical framework. Techniques such as Minimum Redundancy Maximum Relevance (MrMr), Fast Correlation- Based Filter (FCBF), Lasso, Relief, and ANOVA offer various strategies for identifying the most informative features from a dataset. MrMr aims to select features that are highly relevant to the target variable while minimizing redundancy among the features. FCBF, on the other hand, prioritizes features based on their correlation with the target variable and among themselves, ensuring that the selected features provide complementary information. Lasso, a regularization technique, penalizes the inclusion of less important features, thereby enhancing model simplicity and interpretability. Relief focuses on feature weighting based on their ability to discriminate between instances that are close to each other. ANOVA assesses the statistical significance of features in explaining the variance in the target variable. Each of these methods offers unique benefits and can be tailored to the specific characteristics of the dataset and the clinical objectives.

1.4 EXISITING METHODOLOGIES

Existing methodologies for cardiovascular disease detection encompass traditional clinical assessments, invasive diagnostic procedures, and non-invasive imaging techniques. These approaches are increasingly being supplemented by advanced machine learning models and feature selection techniques, such as MrMr, FCBF, Lasso, Relief, and ANOVA, to enhance predictive accuracy and diagnostic efficiency. in smart industries, highlighting their contributions to fault prediction.

- **Random Forest:**

Random Forest is an ensemble learning method that combines multiple decision trees during training and outputs the mode of the classes for classification tasks or the mean prediction for regression tasks. It excels in identifying the most important features for prediction, making it suitable for analyzing complex datasets with numerous variables.

In predictive maintenance, Random Forest is used to predict faults in industrial equipment by combining multiple decision trees. The ensemble approach improves accuracy and generalizability, making it effective for handling diverse datasets with various features. Random Forest identifies the most important features contributing to machine states or faults. This information aids in understanding the critical factors leading to specific predictions, providing insights into the health of industrial equipment. Random Forest is

- **XGBoost:**

XGBoost (Extreme Gradient Boosting) is a scalable and efficient implementation of gradient boosting algorithms. It builds a sequence of decision trees and combines their predictions. It is effective in capturing non-linear relationships in data and is suitable for predictive maintenance scenarios with complex patterns.

XGBoost is employed for anomaly detection in smart industries. It enhances decision trees' performance, making it effective in capturing complex patterns in data from machinery. XGBoost is applied to model the characteristic behavior of critical components in industrial equipment, such as gearboxes and generators. XGBoost is particularly useful for monitoring real-time data and predicting potential faults

- **Logistic Regression:**

Logistic regression is a widely used statistical method for binary classification tasks. In the context of cardiovascular disease detection, it models the probability of a patient having the disease based on various predictor variables. This technique is valued for its simplicity, interpretability, and effectiveness in handling large datasets

- **Gradient boosting:**

Gradient boosting is a powerful machine learning technique that builds an ensemble of decision trees, iteratively improving the model by minimizing prediction errors. It is highly effective for cardiovascular disease detection, offering high accuracy and robustness by combining the strengths of multiple weak learners into a strong predictive model.

- **Support Vector Machines (SVM):**

SVM is a supervised learning algorithm used for classification and regression tasks. It finds an optimal hyperplane in an N-dimensional space that distinctly classifies data points. SVM is applied for binary classification tasks to predict the health state of equipment. It is particularly effective when the relationship between features and outcomes is non-linear, SVM helps classify machinery as healthy or at risk of failure. SVM identifies the optimal decision boundary between different classes, contributing to accurate classification in predictive maintenance

1.5 PROPOSED METHODOLOGY

The proposed methodology for enhancing cardiovascular disease (CVD) detection revolves around a data-driven framework integrating optimal feature selection techniques with robust machine learning classifiers. The framework is designed to process and analyze health data through the following phases: data preprocessing, feature selection, model training, performance evaluation, and interpretation of results.

1.5.1 Data Preprocessing

Before applying machine learning algorithms, raw medical datasets must be preprocessed to ensure quality and consistency. This step involves:

- **Data Cleaning**: Removing duplicate entries, handling missing values through imputation, and eliminating inconsistencies.
- **Normalization/Standardization**: Scaling numerical features to ensure uniformity and improve algorithm performance.
- **Encoding Categorical Variables**: Transforming categorical data (e.g., gender, chest pain type) into numerical format using techniques such as one-hot encoding or label encoding.
- **Splitting the Dataset**: Dividing the data into training and testing sets (commonly using a 70:30 or 80:20 ratio) to validate the model's performance.

1.5.2 Feature Selection Techniques

The next critical step involves applying optimal feature selection methods to reduce dimensionality and enhance model accuracy. The techniques considered in this research include:

- **Minimum Redundancy Maximum Relevance (MrMr)**: Selects features that are most relevant to the target and minimally redundant.
- **Fast Correlation-Based Filter (FCBF)**: Uses correlation measures to filter out irrelevant and redundant features.
- **Lasso (Least Absolute Shrinkage and Selection Operator)**: Regularization method that drives insignificant feature coefficients to zero.
- **Relief Algorithm**: Estimates feature importance based on their ability to distinguish between instances that are near each other.
- **ANOVA (Analysis of Variance)**: Identifies features with statistically significant differences across target classes.

Each technique is applied separately to understand its individual impact on classifier performance.

1.5.3 Classifier Training and Testing

The refined feature sets are used to train multiple machine learning classifiers. The models selected for this study include:

- **Random Forest**
- **XGBoost**
- **Logistic Regression**
- **Gradient Boosting**
- **Support Vector Machine (SVM)**

These models are trained using the selected features and evaluated using the testing dataset. Metrics such as accuracy, precision, recall, and F1-score are computed for each combination of feature selection method and classifier.

1.5.4 Evaluation Metrics

To ensure a comprehensive assessment of each model's performance, several evaluation metrics are used:

- **Confusion Matrix**: Provides insight into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).
- **ROC-AUC Curve**: A graphical representation showing the trade-off between sensitivity and specificity. The area under the curve (AUC) is used as a performance metric.
- **Cross-validation**: 5-fold or 10-fold cross-validation is used to validate model robustness and generalizability.

1.6 IMPLEMENTATION DETAILS

The implementation phase includes the use of tools and libraries widely adopted in data science and healthcare analytics. The programming language used is Python due to its rich ecosystem for machine learning and data processing. Key libraries include:

- **Pandas & NumPy**: For data manipulation and numerical computations.
- **Scikit-learn**: For feature selection techniques, classification models, and evaluation metrics.
- **Matplotlib & Seaborn**: For data visualization and plotting ROC curves, histograms, and heatmaps.
- **XGBoost**: For training and tuning the XGBoost classifier.

Each feature selection technique and classifier combination is executed on the small dataset, followed by visualization and comparative analysis.

1.7 ANALYSIS AND DISCUSSION

The analysis focuses on comparing the performance of classifiers using different feature selection methods. The key aspects include:

- **Effectiveness of Feature Selection**: By examining the number of features selected and their impact on classification accuracy.
- **Model Comparison**: Evaluating how each model performs with each feature selection technique. Random Forest and XGBoost are expected to handle noisy data better, whereas Logistic Regression benefits from fewer, well-selected features.
- **Interpretability**: Simpler models like Logistic Regression and feature importance from Random Forest help in understanding which medical factors are most predictive.
- **ROC-AUC Insights**: AUC values indicate the discriminative ability of each model, where higher AUC suggests better classification performance.
- **Correlation Analysis**: Pearson correlation heatmaps reveal multicollinearity and help justify the elimination of redundant features.

1.8 SIGNIFICANCE OF STUDY

- **Non-invasive Diagnosis**: Machine learning models trained on selected features from routine check-ups can identify at-risk individuals early without invasive tests.
- **Resource Optimization**: By reducing the number of diagnostic tests needed, healthcare providers can allocate resources more efficiently.
- **Personalized Treatment**: Identifying key risk factors can help tailor treatment plans for individual patients.
- **Future Integration**: The models can be embedded into clinical decision support systems (CDSS) for real-time patient monitoring and diagnostics.

1.8.1 Advancement of Non-Invasive Diagnostics

- The use of machine learning (ML) and optimal feature selection techniques marks a significant advancement in non-invasive diagnostic tools for cardiovascular disease (CVD). Traditional diagnostic approaches often involve invasive procedures such as angiography or stress testing, which can be expensive, time-consuming, and associated with patient discomfort or risk. In contrast, ML-based methods utilize existing clinical data—such as electrocardiograms (ECGs), blood tests, and patient history—to develop predictive models without the need for physical intervention.
- By analyzing patterns within patient data, ML algorithms can detect early signs of CVD that may not be evident to clinicians using conventional diagnostic criteria. Feature selection enhances this process by identifying the most informative attributes, improving

model interpretability and reducing noise. This makes it feasible to deploy ML tools in a clinical setting, potentially enabling rapid screening and early intervention for at-risk individuals.

- Moreover, non-invasive diagnostics are particularly valuable in remote or resource- limited settings, where access to sophisticated imaging equipment or specialized medical personnel is constrained. By integrating ML models into portable or wearable devices, healthcare providers can conduct real-time monitoring and risk assessments, broadening access to quality care.
- In summary, the incorporation of machine learning and optimal feature selection into non- invasive diagnostics not only reduces patient risk but also enables scalable, accessible, and cost-effective CVD detection, aligning well with global health priorities for early disease prevention and management.

1.8.2 Enhanced Predictive Accuracy and Reliability

- One of the primary objectives of incorporating machine learning (ML) in cardiovascular disease (CVD) detection is to enhance the predictive accuracy and reliability of diagnostic models. Traditional statistical methods often struggle with high-dimensional datasets, common in medical diagnostics, leading to reduced model performance or overfitting. Feature selection techniques such as Lasso, Relief, ANOVA, and MrMr play a pivotal role in addressing this issue by extracting the most relevant and non-redundant variables from the dataset.
- By focusing only on significant predictors, the dimensionality of the dataset is reduced, which leads to the construction of more generalizable models with better accuracy across unseen data. This streamlined data representation also improves model interpretability, which is crucial in clinical decision-making. Enhanced predictive reliability ensures that high-risk patients are accurately identified, minimizing both false positives and false negatives. This translates into more efficient allocation of medical resources and improved patient outcomes.
- Additionally, ensemble learning methods such as Random Forests and XGBoost further strengthen predictive performance by reducing the variance and bias associated with individual classifiers. These models aggregate predictions from multiple learners, resulting in more robust and consistent outputs. The integration of these advanced models with feature selection methodologies provides a powerful framework for CVD detection.
- Ultimately, the enhancement of predictive accuracy and reliability directly contributes to improved healthcare delivery, as it supports evidence-based decisions, early interventions,

and targeted treatment plans, all of which are essential in managing and preventing cardiovascular events.

1.8.3 Optimization of Clinical Decision Support Systems

- Clinical Decision Support Systems (CDSS) have become integral in modern healthcare for aiding physicians in diagnosis, treatment planning, and patient monitoring. The application of machine learning (ML) and optimal feature selection significantly contributes to the optimization of CDSS, particularly in the context of cardiovascular disease (CVD). These technologies enable the systems to process vast datasets and extract meaningful insights that may not be readily observable by clinicians.
- Feature selection methods such as FCBF, Relief, and ANOVA refine CDSS by identifying the most predictive attributes, thereby reducing data redundancy and noise. This results in faster processing times and more streamlined decision paths within the system. Additionally, incorporating ML algorithms ensures that the CDSS evolves with time and improves as more data becomes available, making it adaptive and context-aware.
- An optimized CDSS can assist in early CVD detection, risk stratification, and personalized treatment recommendations by integrating patient-specific data with medical knowledge. This improves diagnostic accuracy and reduces variability in clinical decisions, especially in complex or ambiguous cases. Moreover, it enhances the efficiency of healthcare delivery by automating routine analyses, thereby allowing healthcare professionals to focus on patient care.
- Furthermore, the use of transparent ML models with interpretable feature selection provides clinicians with justifications for specific recommendations, fostering trust in the system. Overall, the integration of ML-driven optimization into CDSS is a critical step toward precision medicine, ensuring that healthcare is more predictive, preventive, and patient-centered.

1.8.4 Cost-Effective Screening Solutions

- The implementation of machine learning (ML) and optimal feature selection in cardiovascular disease (CVD) detection presents a compelling case for cost-effective screening solutions, especially in resource-constrained environments. Traditional diagnostic procedures such as echocardiography, angiography, or advanced imaging techniques are expensive and often require significant infrastructure and trained specialists. In contrast, ML-based systems leverage readily available patient data—such as age, blood pressure, cholesterol levels, and lifestyle factors—to predict CVD risk with high accuracy.

- Feature selection techniques further enhance the cost-efficiency of these systems by focusing on the most informative and relevant variables, thereby reducing computational complexity and unnecessary diagnostic tests. This streamlining minimizes operational expenses while maintaining or even improving diagnostic precision. For example, using only the top-ranked features can lead to robust predictions with fewer diagnostic inputs, making it feasible to implement CVD screening in community clinics and primary care centers.
- Moreover, integrating ML models into mobile or cloud-based platforms allows widespread deployment with minimal infrastructure. Healthcare providers in rural or underserved areas can use these tools for early detection, reducing the burden on tertiary hospitals and preventing costly late-stage interventions. Preventive care driven by accurate early detection not only improves patient outcomes but also reduces long-term healthcare expenditures.
- Overall, by reducing dependency on high-end diagnostics and enabling scalable deployment, ML and feature selection offer a transformative shift toward affordable, proactive CVD screening solutions that are both effective and economically sustainable.

1.8.5 Empowering Healthcare in Resource-Limited Regions

- Machine learning (ML) and optimal feature selection have a transformative role in empowering healthcare delivery in resource-limited regions, where access to advanced medical facilities and specialized personnel is often scarce. Cardiovascular diseases (CVD) are a major cause of mortality in low- and middle-income countries, where early detection and timely intervention are critical but frequently unattainable due to infrastructural and financial constraints.
- By utilizing non-invasive, easily obtainable patient data—such as heart rate, blood pressure, lifestyle habits, and basic laboratory tests—ML models can predict CVD risk with high accuracy. Feature selection techniques help distill this data to its most essential components, ensuring that models remain accurate while being lightweight and computationally efficient. These refined models can be deployed on low-resource platforms such as mobile phones or handheld devices, enabling frontline healthcare workers to conduct screenings without needing expensive equipment.

Furthermore, the democratization of ML tools through open-source platforms and cloud integration makes it possible to scale diagnostic capabilities across remote and rural regions. Community health programs can leverage these tools to perform mass screenings and prioritize high-risk individuals for further medical attention.