|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Title**  *(Name of the journal, author and publication details)* | **Methodology**  *(Provide a Summary of key studies and their findings)* | **Identification of gaps and limitations.**  *(Identify the limitations of the Research Paper)* |
| 1. | **Effective Heart Disease Prediction Using Machine**  **Learning Techniques**  **Journal: Algorithms**  **Authors: Chintan M. Bhatt, Parth Patel, Tarang Ghetia, Pier Luigi Mazzeo**  **Publication Details: Algorithms 2023, 16, 88. Published on February 6, 2023** | * **Approach: Utilized machine learning algorithms including random forest, decision tree, multilayer perceptron, and XGBoost for heart disease prediction.** * **Data: Analyzed a dataset of 70,000 instances from Kaggle.** * **Accuracy: Achieved up to 87.28% accuracy with multilayer perceptron.** | * Feature Selection: Insufficient discussion on the ranking and optimality of features * Model Limitation: Limited generalizability for smaller datasets; dependent on data quality and preprocessing methods. |
| 2. | **Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection**  **Journal: IEEE Access Authors: Tahseen Ullah, Syed Irfan Ullah, Khalil Ullah, Muhammad Ishaq, Ahmad Khan, Yazeed Yasin Ghadi, Abdulmohsen Algarni Publication Details: Received 3 December 2023, accepted 13 January 2024, published 30 January 2024, current version 5 February 2024** | * **Data Collection: Collecting ECG signals.** * **Feature Extraction: Extracting features from ECG signals.** * **Feature Selection: Using FCBF, MrMr, Relief, and PSO-optimization for selecting optimal features.** * **Machine Learning Models: Training Extra Tree and Random Forest classifiers on selected features.** | * Generalizability: Limited generalizability due to small dataset size. * Feature Selection: Insufficient discussion on the ranking and optimality of features. * Comparison: Lack of comparison with state-of-the-art methods. * Ethical Considerations: Ethical issues in healthcare research not thoroughly explored. |
| 3. | **Detection of Cardiovascular Diseases Using Machine Learning Models**  **Journal: Diagnostics**  **Authors: Adedayo Ogunpola, Faisal Saeed, Shadi Basurra, Abdullah M. Albarrak, Sultan Noman Qasem**  **Publication Details: 21 December 2023, Accepted: 25 December 2023, Published: 8 January 2024.** | * **Data Collection: Used coronary artery X-ray angiography images and heart failure patient records.** * **Preprocessing: Applied oversampling, feature scaling, normalization, and dimensionality reduction.** * **Feature Extraction: Extracted features like color, diameter, shape, age, anemia, and high blood pressure.** * **Machine Learning Models: Implemented KNN, SVM, Logistic Regression, CNN, Gradient Boost, XGBoost, and Random Forest.** * **Evaluation: Achieved high accuracy and precision with the XGBoost model.** | * Imbalanced Datasets: Challenge of handling imbalanced datasets. * Dataset Variability: Variability in datasets complicates predictions. * Generalizability: Ensuring model generalizability to diverse populations. * Model Comparison: Need for comprehensive comparison with state-of-the-art methods.data. |
| 4 | **Paper Title: Heart Disease Detection Using Machine Learning Models**  **Journal: Procedia Computer Science**  **Authors: Amrit Singh, Harisankar Mahapatra, Anil Kumar Biswal, Madhumita Mahapatra, Debabrata Singh, Milan Samantaray**  **Publication Details: Procedia Computer Science, Volume 235, Pages 937-947, Published by Elsevier B.V.**  **Publication Date: 2024** | * Data Collection: Utilized the Framingham Heart Study dataset. * Preprocessing: Managed missing data, outliers, and balanced class distribution. * Feature Selection: Applied Principal Component Analysis (PCA). * Machine Learning Models: Implemented Decision Tree, Random Forest, and Support Vector Machine (SVM). * Evaluation: Compared accuracy using metrics like AUC-ROC, accuracy, sensitivity, and specificity | * **Data Collection: Utilized the Framingham Heart Study dataset.** * **Preprocessing: Managed missing data, outliers, and balanced class distribution.** * **Feature Selection: Applied Principal Component Analysis (PCA).** * **Machine Learning Models: Implemented Decision Tree, Random Forest, and Support Vector Machine (SVM).** * **Evaluation: Compared accuracy using metrics like AUC-ROC, accuracy, sensitivity, and specificity** |
| 5 | Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset  Authors: Khandaker Mohammad Mohi Uddin, Rokaiya Ripa, Nilufar Yeasmin, Nitish Biswas, Samrat Kumar Dey  Published in Intelligence-Based Medicine, Volume 7, 2023, Article 100100  Available online 29 April 2023 | This study investigates the use of supervised machine learning (ML) algorithms for early and accurate disease detection. The methodology involves evaluating multiple supervised ML models such as KNN, NB, DT, CNN, SVM, and LR for heart, kidney, breast, and brain diseases. The analysis includes comparing the performance of these algorithms to determine the most effective models for each disease​ | * **Model Details**: Lack of details on neural network parameters. * **Dataset Quality**: Challenges in managing the quality and selection of datasets. * **Feature Selection**: Difficult and computationally intensive. * **Preprocessing**: Hindered ML model performance by potentially reducing data quality. * **Small Dataset Size**: Constrained the learning models from achieving higher accuracy and precision. |
| 6 | **An automatic diagnostic model for the detection and classification of cardiovascular diseases based on swarm intelligence technique  Journal: Heliyon Authors: C. Venkatesh, B.V. V. S. Prasad, Mudassir Khan, J. Chinna Babu, M. Venkata Dasu Publication Details: Heliyon 10 (2024) e25574, available online 5 February 2024.** | * Objective: Develop an automatic diagnostic model using deep learning and swarm intelligence to detect and classify cardiovascular diseases. * Techniques Used: The model integrates convolutional neural networks (CNN) with particle swarm optimization (PSO). * Data Sources: ECG signals from private hospital patients and the MIT-BIH Arrhythmia database. * Findings: Achieved 99.58% accuracy in predicting cardiovascular diseases, outperforming previous models that reported accuracies between 83.52% and 96.72%. | * **Challenges: Issues with the large dimension and class imbalance of clinical datasets.** * **Generalizability: Limited validation on specific datasets may affect generalizability to other clinical data or patient populations.** * **Computational Resources: High computational demands may limit practicality in resource-constrained settings.** * **Comparative Studies: Need for further studies comparing advanced methods with traditional models.** * **Metrics Evaluation: Lacks detailed analysis of trade-offs between different performance metrics.** |
| 7 | **Recent advancements using machine learning & deep learning approaches for diabetes detection**  **Journal: e-Prime - Advances in Electrical Engineering, Electronics and Energy**  **Authors: Neha Katiyar, Hardeo Kumar Thakur, Anindya Ghatak**  **Publication Details: e-Prime - Advances in Electrical Engineering, Electronics and Energy 9 (2024) 100661, available online 28 June 2024.** | * Objective: To provide a comprehensive review of machine learning (ML) and deep learning (DL) techniques applied to the early detection and management of Diabetes Mellitus. * Techniques Reviewed: Various ML and DL techniques, including Support Vector Machines (SVM), Regression, k-nearest Neighbours (KNN), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). * Data Sources: Datasets such as Pima Indian Diabetes Dataset (PIDD) and others. * Findings: The review covers the application of these techniques in different aspects of diabetes detection and management, including classification, detection, and prediction. It highlights advancements and performance metrics achieved by different models. | * **Data Quality: Issues with the quality and completeness of datasets, including imbalance and the need for more diverse datasets.** * **Generalizability: Limited generalizability of findings due to the reliance on specific datasets.** * **Model Complexity: Complexity of models which may hinder their practical implementation in real-world clinical settings.** * **Computational Resources: High computational requirements for training advanced models, limiting their use in resource-constrained environments.** * **Validation: Need for more extensive validation studies to confirm the robustness of models across different populations and settings.** |
| 8 | **The future of cardiovascular disease prevention with**  **machine learning and digital health technology**  **Journal: Heliyon**  **Authors: C. Venkatesh, B.V. V. S. Prasad, Mudassir Khan, J. Chinna Babu, M. Venkata Dasu**  **Publication Details: Heliyon, Volume 10, 2024, e25574. Available online 5 February 2024. Published by Elsevier Ltd. This is an open access article under the CC BY license.** | * CVD Detection Using PSO: Applied PSO with ECG data for accurate cardiovascular disease classification. * ML for CVD Prevention: Reviewed ML/DL approaches showing superior accuracy over traditional risk tools in cardiovascular disease prevention. * ML Models for Diabetes: Validated various ML models, including fuzzy logic and non-invasive methods, to enhance diabetes diagnosis and prediction accuracy. | * **Data Availability: Limited by fragmented data silos and lack of external validation for most healthcare ML models​** * **Model Development: Insufficient volume, diversity, and reproducibility of training data hinder ML model effectiveness in clinical contexts​** |
| 9 | **Heart Disease Detection Using Machine Learning Models**  **Journal: Procedia Computer Science**  **Authors: Amrit Singh, Harisankar Mahapatra, Anil Kumar Biswal, Madhumita Mahapatra, Debabrata Singh, Milan Samantaray**  **Publication Details: Procedia Computer Science 235 (2024) 937–947, Published by ELSEVIER B.V., Open Access under CC BY-NC-ND license. Available online at ScienceDirect.** | * Hybrid ML algorithm using PSO and SVM improved classification accuracy. * Ensemble model (logistic regression + naive Bayes) outperformed existing models. * Systematic review highlighting challenges with imbalanced datasets in ML. * Utilized various ML techniques to predict cardiovascular diseases from high-dimensional data. * Developed a Stacking Ensemble Learner using XgBoost for emergency readmission prediction. | * **Overfitting: Existing models often overfit the training data.** * **Data Imbalance: Imbalanced datasets can skew model performance.** * **Complexity and Diversity: Diverse heart disease manifestations complicate diagnosis.** * **Dataset Requirements: Need for large, diverse datasets for robust models.** * **Evaluation Metrics: Comprehensive performance evaluation is crucial.** |
| 10 | **ECG-based heartbeat classifier for early detection of cardiovascular disease, hypertrophic cardiomyopathy (HCM)  Journal: Informatics in Medicine Unlocked**  **Authors: R. Hagan et al.**  **Publication Details: Informatics in Medicine Unlocked, Volume 24, 2021, Article 100606​** | * Data Collection: ECG data from HCM patients using standard 10-s, 12 lead ECG signals. Heartbeats from non-HCM cardiovascular patients serve as controls. * Algorithms: Random Forest, Support Vector Machine (SVM), and Logistic Regression classifiers. Hyperparameter optimization was performed using k-fold cross-validation. * Metrics: Performance evaluated using accuracy, precision, and F1 score. | * **Dataset Specificity: Focus on HCM patients may limit generalizability to other heart diseases.** * **Limited Scope: Analysis limited to two datasets, which may not capture the full variability of heart disease presentations​** |
| 11 | **Title: A Survey on Machine Learning in Congenital Heart Disease Diagnosis: State of the Art and Future Directions**  **Journal Name: Technologies**  **Author: Not specified**  **Publication Details: 2024, Volume 12, Page 4** | * Proposed a method for identifying pulmonary hypertension associated with CHD using time-frequency domain analysis and ML, achieving a precision level of 91.6%​ * explored AI in assessing CHD, focusing on ML for image analysis, risk estimation, and detection, highlighting AI's potential in advancing cardiac care​ * Introduced a lightweight CNN architecture for classifying radiation-induced liver disease, achieving a classification accuracy of 93.1%, showcasing ML's versatility in medical image analysis​ | * **IoMT Data Utilization: Need for more research on using IoMT data for cardiovascular diagnosis.** * **LSTM and Attention Systems: Further exploration needed for cardiovascular disease detection.** * **CNN-BiLSTM-AM: Lack of research on combined use for precise diagnostics.** * **Model Assessment and Comparison: Need for thorough comparison using Heart Disease UCI and Cardiovascular Disease Dataset.** * **Congestive Heart Failure Detection: Improvement needed in accuracy, reliability, and feature extraction efficiency.** |
| 12 | **Deep Learning Approach for Automatic Cardiovascular Disease Prediction Employing ECG Signals**  **Journal Name: Computer Modeling in Engineering & Sciences**  **Authors: Muhammad Tayyeb, Muhammad Umer, Khaled Alnowaiser, Saima Sadiq, Ala’ Abdulmajid Eshmawi, Rizwan Majeed, Abdullah Mohamed, Houbing Song, Imran Ashraf**  **Publication Details: Published online June 13, 2023** | * Study Focus: This study proposes a multilayer perceptron (MLP) model for heart disease prediction using ECG data. * Data: ECG dataset with averaged signals using a window size of 10. * Comparison Models: Several deep learning (CNN, RNN, LSTM) and machine learning models (logistic regression, extra tree classifier, stochastic gradient descent, random forest, Gaussian Naive Bayes) were used. * Validation: K-fold cross-validation was applied. * Results: The MLP-based model achieved a 94.40% accuracy score. | * **Model Complexity: While the MLP model reduces computational complexity, further reduction and optimization may be needed for practical deployment.** * **Feature Selection: The study highlights the challenge of manual feature extraction and the risk of models overfitting or underfitting due to inappropriate feature selection.** * **Real-world Application: The study's promising results indicate potential for real-world application, but further validation in diverse clinical settings is necessary.** * **Comparison Scope: The study could benefit from a broader comparison with additional models and datasets to establish the robustness and generalizability of the proposed approach.** * **Limited Dataset: The ECG dataset used may not capture all possible variations in cardiovascular disease presentations, suggesting a need for larger and more varied datasets in future research.** |
| 13 | **A Novel Machine Learning-Based Prediction Method for Early Detection and Diagnosis of Congenital Heart Disease Using ECG Signal Processing**  **Authors : Prabu Pachiyannan 1,Musleh Alsulami 2,Deafallah Alsadie 3,Abdul Khader Jilani Saudagar ,Mohammed AlKhathami andRamesh Chandra Poonia 5**  **Journal: Technologies**  **Publication Details: Technologies 2024, 12, 4, pages 2-23** | * The study used machine learning algorithms KNN, Logistic Regression, and Random Forest to predict heart disease. * Data was collected, preprocessed, and then classified using the mentioned algorithms. * Performance was evaluated based on accuracy and other metrics. * The Random Forest algorithm was specifically noted for its superior accuracy in disease prediction. * Comparisons were made between different models to demonstrate the efficiency of the proposed method​ | * **Limited size of the dataset used for training the models.** * **Models need enhancement for scalability and handling multi-parameter inputs.** * **The study mainly focused on the symptoms without considering the complete medical history of patients.** * **Some methodologies have issues like overfitting and are not suitable for multi-parameter scenarios** |
| 14 | **Disease Prediction Using Machine Learning**  **Journal: International Journal of Scientific Research in Science and Technology**  **Authors: Gaurav Shilimkar, Shivam Pisal**  Publication Details: Volume 8, Issue 3, Pages 551-555, May-June 2021 | * Data Collection: Data was sourced from a local hospital in Pune. * Data Transformation: Raw data was cleaned and converted into a format suitable for analysis. * Feature Extraction: Relevant features were extracted to improve the model's performance. * Model Training: The decision tree algorithm was trained on the dataset to create a prediction model. * Implementation: The system had two components—admin for training and user for inputting symptoms and obtaining predictions. * Evaluation: The model's efficacy was assessed using real-life clinical data. | * **Limited Scope: The system currently covers only general and commonly occurring diseases.** * **Data Quality: Predictions can be weakened by incomplete or imperfect data.** * **Regional Bias: The system may not accurately predict diseases that are specific to regions different from where the training data was sourced.** * **Future Expansion Needed: The system does not currently include serious illnesses like various types of cancer, which would benefit from early detection.** |
| 15 | **Disease Prediction using Machine Learning**  **Journal: International Journal for Research in Applied Science & Engineering Technology (IJRASET)**  **Authors: Kriti Gandhi, Mansi Mittal, Neha Gupta, Shafali Dhall**  **Publication Details: Volume 8, Issue VI, June 2020** | * K-Nearest Neighbours (KNN): Non-parametric and lazy learning algorithm. * Logistic Regression: Regression analysis for binary dependent variables. * Decision Tree: Supervised learning algorithm used for classification. * Naive Bayes: Classification technique based on Bayes' Theorem. * Linear Discriminant Analysis (LDA): Dimensionality reduction algorithm. * Random Forest: Ensemble learning method for classification. | * **Dataset Limitations: The dataset used was limited to 132 symptoms and 40 diseases, which might not cover all possible medical conditions.** * **Overfitting Concerns: Initial model fitting might lead to overfitting, which was addressed by reducing the number of features.** * **Algorithm Performance Variability: Different algorithms showed varied performance, with Logistic Regression and KNN performing better, while LDA showed the lowest performance.** * **Generalization Issues: The model's accuracy might be limited to the specific dataset and may not generalize well to different datasets or real-world scenarios.** |
| 16 | **Effective Heart Disease Prediction Using Machine Learning Techniques**  **Journal: Algorithms**  **Authors: Chintan M. Bhatt, Parth Patel, Tarang Ghetia, Pier Luigi Mazzeo**  **Publication Details: Algorithms 2023, 16, 88. Published on February 6, 2023** | * Approach: Utilized machine learning algorithms including random forest, decision tree, multilayer perceptron, and XGBoost for heart disease prediction. * Data: Analyzed a dataset of 70,000 instances from Kaggle. * Accuracy: Achieved up to 87.28% accuracy with multilayer perceptron. | * **Dataset Size: The study mitigated overfitting by using a large dataset, unlike prior research with smaller datasets.** * **Model Limitation: Limited generalizability for smaller datasets; dependent on data quality and preprocessing methods.** |



**LIMITATIONS(base paper):-**

* **Feature Selection**: Insufficient discussion on the ranking and optimality of features
* **Model Limitation**: Limited generalizability for smaller datasets; dependent on data quality and preprocessing methods
* **Risk of Overfitting**: Improper feature selection can lead to overfitting, where the model performs well on training data but poorly on unseen data. This happens when selected features capture noise rather than underlying patterns.