IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Intelligent and Efficient Methodlogy for Heart Disease Detection using Autoencoder

Dr. V. Gowri
Dept. Of Computer Science and Engineering
SRMIST,Ramapuram
Chennai, India

Ankit Kumar
Dept. Of Computer Science and Engineering
SRMIST,Ramapuram
Chennai, India

Balasubramaniam
Dept. Of Computer Science and Engineering
SRMIST,Ramapuram
Chennai, India

Deepak Mishra
Dept.Of Computer Science Engineering
SRMIST,Ramapuram
Chennai, India

Abstract— Artificial Intelligence (AI) is revolutionizing healthcare by leveraging data and advanced analytics, particularly in medical diagnosis where the stakes are high. However, concerns about transparency, explainability, and bias persist. This study focuses on using Neural Networks to analyze a Heart diseases dataset, starting with understanding feature correlations and applying feature selection techniques for optimal detection of heart disease. By experimenting with deep learning classification models using both complete and reduced feature subsets, the study shows that relevant features significantly impact classification accuracy. Even with fewer features, models perform better and train faster, highlighting the potential of AI in improving healthcare outcomes.

INTRODUCTION

The modern hectic lifestyle often leads to unhealthy habits, contributing to anxiety and depression. This can lead to increased smoking, drinking, and drug use, which are major factors in various dangerous diseases like cardiovascular diseases and cancer. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) have the highest mortality rates globally, accounting for about 31% of all deaths. Early detection of these diseases is crucial for taking precautionary measures before they become severe. CVDs encompass conditions affecting the heart or blood vessels, including Coronary Heart Disease, Stroke/Transient Ischemic Attack (TIA/MiniStroke), Peripheral Arterial Disease, and Aortic Disease. While the exact cause of CVDs remains unknown, risk factors such as high blood pressure, smoking, diabetes, body mass index (BMI), cholesterol levels, age, and family history play significant roles. The challenge lies in accurately predicting these diseases early to reduce mortality rates through effective treatments and interventions.

In the realm of ECG classification, various methods exist, from feature-based models to deep learning approaches. The advancement of healthcare technology has led to the development of Healthcare Cyber-Physical Systems (H-CPS), integrating electronic health records (EHR) with artificial intelligence (AI) for smarter healthcare management. This includes monitoring patient conditions through the Internet of Medical Things (IoMT) and wearable sensors, enabling efficient disease diagnosis and remote monitoring, particularly crucial for cardiovascular diseases.

The early detection of CVDs is essential for implementing preventive measures and reducing premature mortality. Past efforts have explored machine learning models for automated CVD detection, paving the way for intelligent preventive healthcare systems integrated with smart healthcare technologies. Portable medical devices and non-invasive monitoring tools have simplified everyday health monitoring, aiding in the early diagnosis and management of conditions like arrhythmias, heart rate abnormalities, and chronic heart failure through ECG monitoring and data analysis from Electronic Health Records (EHR).

Challenges in HSI classification include limited ground-truth data, high dimensionality, low spatial quality, and spectral variability. Deep learning-based classifiers, with their automatic feature learning and high precision

I. Literature Survey

Literature Survey 1

Title: An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases

Authors: Aqsa Rahim, Yawar Rasheed, Farooque Azam, Muhammad Waseem Anwar, Muhammad Abdul Rahim, Abdul Wahab Muzaffar

Published Year: 2021

Efficiency: Simple to understand and interpret, delivers information on the state of operation, boosts performance

Drawbacks: High complexity of installation and maintenance, opportunistic and uncontrollable, computation burden may limit further application in real scenarios

Description: This framework, MaLCaDD, addresses the challenge of predicting cardiovascular diseases by proposing an ensemble of machine learning techniques. It handles missing values, data imbalance, and performs feature selection before utilizing logistic regression and K-Nearest Neighbor classifiers for prediction. Achieves high accuracies on benchmark datasets.

Literature Survey 2

Title: Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques

Authors: Pronab Ghosh, Sami Azam, Mirjam Jonkman, Asif Karim, F. M. Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, Friso De Boer

Published Year: 2021

Efficiency: Fast and efficient, achieves sub-optimal performance, effective for distributed optimization

Drawbacks: Difficult and less commonly used, may take a huge time and economic cost to construct, high level of communication and computation overheads

Description: This approach focuses on predicting cardiovascular diseases using efficient data collection, preprocessing, and transformation methods along with Relief and LASSO feature selection techniques. It integrates traditional classifiers with bagging and boosting methods to improve prediction accuracy.

Literature Survey 3

Title: Multi-Compartment Spatially-Derived Radiomics From Optical Coherence Tomography Predict Anti-VEGF Treatment Durability in Macular Edema Secondary to Retinal Vascular Disease: Preliminary Findings

Authors: Sudeshna Sil Kar, Duriye Damla Sevgi, Vincent Dong, Sunil K. Srivastava, Anant Madabhushi, Justis P. Ehlers

Published Year: 2021

Efficiency: Achieves sub-optimal performance, reduces resources used for processing, tolerates variations

Drawbacks: High complexity of installation and maintenance, hard to maintain, poor application performance

Description: This study explores the use of radiomics features extracted from different spatial compartments of OCT images to predict treatment durability in retinal vascular diseases. It identifies texture-based radiomics features that discriminate between patients who tolerate extended treatment intervals and those who require more frequent dosing.

Literature Survey 4

Title: An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier

Authors: Mohammad Ayoub Khan

Published Year: 2020

Efficiency: Corresponding time cost is greatly reduced, reduces consumption of hardware resources, offers increased flexibility

Drawbacks: Computation burden may limit further application in real scenarios, have not been investigated thoroughly, may take a huge time and economic cost to construct

Description: This framework utilizes IoT technology and a Modified Deep Convolutional Neural Network (MDCNN) for heart disease prediction based on data from wearable devices. The proposed method outperforms existing classifiers, showcasing potential for remote health monitoring.

Table 1: Advantages and Disadvantages of Existing Model

S_No:	Advantages	Disadvantages
1	Enhanced feature extraction for hyperspectral data	Not explicitly addressing data imbalance
2	Spatial-spectral fusion improves classification accuracy	Complexity in integrating architectures
3	Addresses limited training samples; effective clustering	May require significant preprocessing for clustering
4	Efficient multisource domain transfer learning; unified space	Dependency on source domain data
5	retention capturing and cardiac features	Potential increased computational complexity

II. Existing Methodology

Learning Hidden Features: One of the primary advantages of autoencoder decoders is their ability to learn important hidden features present in the input data. By training the autoencoder to reconstruct the input data from a compressed representation (latent space), the decoder learns to extract and represent the most salient features of the data. This process helps in reducing the reconstruction error, as the decoder strives to accurately reconstruct the original input using only the learned features.

Dimensionality Reduction: Autoencoder decoders facilitate dimensionality reduction by compressing the size of the input data into a smaller representation. This smaller representation, often referred to as the latent space or encoding, contains the essential features necessary for reconstruction. By reducing the dimensionality of the data, autoencoders enable more efficient storage and processing of information, especially in scenarios where large datasets or high-dimensional inputs are involved.

Efficient Deep Learning Model Creation: Autoencoder decoders play a crucial role in making the creation of deep learning models much more efficient. By learning a compact representation of the input data, autoencoders provide a means to preprocess and extract meaningful features from raw data, which can then be fed into downstream neural network architectures for tasks such as classification, regression, or clustering. This hierarchical approach to feature learning not only enhances the efficiency of deep learning models but also improves their performance by focusing on relevant features.

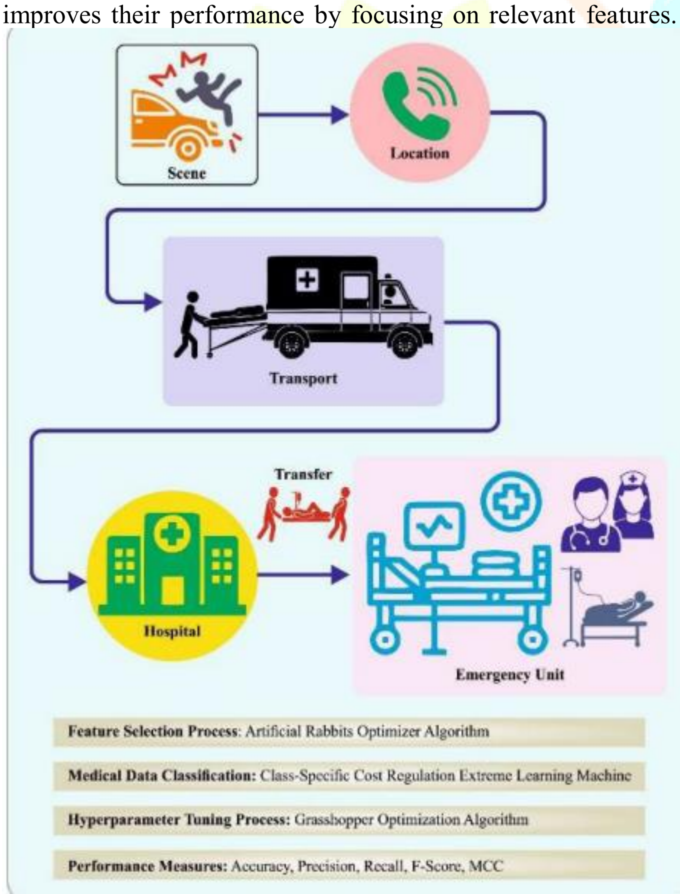


FIGURE 1. Workflow of AUTOENCODER approach.

III. Proposed Methodology

In this study, we propose a novel approach, AUTOENCODER, for medical data classification and monitoring in the Emergency Departments (EDs) of hospitals in the Kingdom of Saudi Arabia (KSA). AUTOENCODER facilitates the monitoring and tracking of patient visit data, treatment provided, and Length of Stay (LOS). The technique involves several sub-processes for medical data classification, including min-max normalization, ARO-based feature subset selection, CSCR-ELM classification, and GOA-based parameter tuning. The workflow of the AUTOENCODER approach is illustrated in Figure 1.

A. DATA PRE-PROCESSING**

We employed min-max normalization to scale the data to unit variance, facilitating the computation of similarity between data points. Min-max normalization ensures that features are scaled to comparable ranges.

B. ALGORITHMIC DESIGN OF ARO-BASED FEATURE SELECTION APPROACH

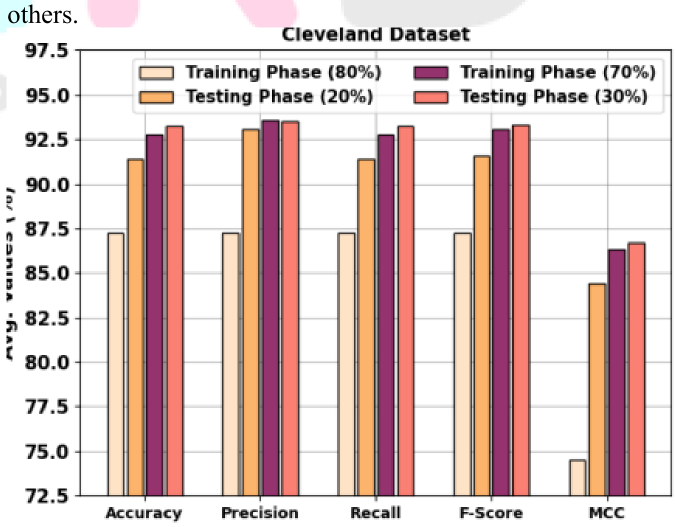
The ARO algorithm is employed in this study to select optimal features. Inspired by the survival strategy of rabbits in nature, the ARO technique is utilized. Rabbits exhibit detour foraging behavior, venturing far from their nests for food, and engaging in random hiding to evade predators. The decision to engage in random hiding or detour foraging depends on their energy levels. When energy is low, they hide randomly nearby; when energy is sufficient, they forage far from their nest.

1) ENERGY SHRINK (SWITCH BETWEEN EXPLOITATION AND EXPLORATION)

The rabbit's decision to engage in detour foraging or random hiding is determined by its energy level. Energy factor $A(t)$ is calculated, determining the action the rabbit takes. If $A(t)$ is less than or equal to 1, random hiding is performed; if $A(t)$ is greater than 1, detour foraging is performed.

2) DETOUR FORAGING (EXPLORATION)

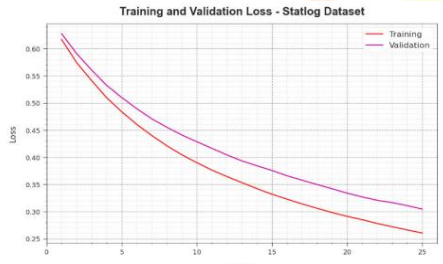
During detour foraging, rabbits search for food far from their nests to safeguard against predators. This exploration process involves randomly searching for food based on the positions of others.



6. SoftMax Classification: The final layer employs a softmax function to generate a probability distribution representing class membership probabilities for each pixel.

V. Results And Disscussions

The majority of the outliers in the data are considered as noise which adds no value to the significance of data and negatively affects the performance of the model. Raveendrababu et al. have demonstrated that removing the outliers from the data helps in achieving the improved results and proposed various methodologies for outlier removal. Our proposed framework uses Boxplot for the removal of outliers. Boxplots are used to give five numbers summary i.e. minimum, first quartile (Q1), median (Q2), third quartile (Q3) and maximum. When these five points are plotted it forms a box like structure and any point which is lying outside this box is considered as an outlier. After removing the outliers, our proposed framework checks for the missing values in the data. Missing values in the data are very common which can exist because of the faulty instrument or human error . If there are missing value in the data then model cannot be efficiently trained on that data due to the reduced number of training samples. This affects the accuracy of the model. That is why our proposed framework replaces the missing values by the mean of all the values of the respective attribute. Mean substitution is one of the most common methods used by various researchers (e.g., Dodeen et al) for replacing the missing values. This helps to retain training data without adding further information to the dataset which reduces the chances of overfitting. After resolving the issues of outliers (via boxplot technique) and missing values (via mean replacement technique), the final step of preprocessing in our proposed framework is to resolve the problem of imbalanced class.



VI. CONCLUSION

In this study, a novel hyperspectral image classification method is introduced, leveraging dense pyramidal convolution and multi-feature fusion to enhance the extraction and utilization of spatial and spectral information from hyperspectral images with limited sample sizes. The methodology employs two distinct branches: a spatial branch and a spectral branch. Within the spatial branch, dense pyramidal convolution layers are utilized alongside non-local blocks to extract both local and global spatial features from the image samples. Concurrently, the spectral branch employs a dense pyramidal convolution module to capture spectral features. Subsequently, the extracted spatial and spectral features are fused using fully connected layers to yield the final classification results.

References

M. A. Malik. (,). Internet of Things Healthcare Market, Allied , Online, . Available: Market Research. Accessed: Oct. , , .

X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, A hybrid classication system for heart disease diagnosis based on the Art. no. , .

R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby, Braunwalds Heart Disease E-Book: A Textbook of Cardiovascular Medicine. Amsterdam, The Netherlands: Elsevier, , .

O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, An integrated decision support system based on ANN and Fuzzy_AHP for Feb. , .

N. Jankowski and V. Kadirkamanathan, Statistical control of RBF- like networks for classication, in Proc. Int. Conf. Artif. Neural Netw.

S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, Use of kernel func- tions in articial immune systems for the nonlinear classication prob- Jul. , .

P. J. Fitzgerald, T. A. Ports, and P. G. Yock, Contribution of localized calcium deposits to dissection after angioplasty. An observational study Jul. , .

M. Gargesha, R. Shalev, D. Prabhu, K. Tanaka, A. M. Rollins, M. Costa, H. G. Bezerra, and D. L. Wilson, Parameter estimation of atheroscle- rotic tissue optical properties from three-dimensional intravascular optiArt. no. , doi: , ., /, .JMI., ., ., .

, J. Lee, D. Prabhu, C. Kolluru, Y. Gharaibeh, V. N. Zimin, H. G. Bezerra, and D. L. Wilson, Automated plaque characterization using deep learn- ing on coronary intravascular optical coherence tomographic images, , ., /BOE., ., . P. Tait, Introduction to Radar Target Recognit., vol. 18, 2005.

A. Abdolmana, L. Duong, N. Dahdah, I. R. Adib, and F. Cheriet, Char- acterization of coronary artery pathological formations from OCT imaging Sep. , doi: , ., /BOE., ., .

M. Liu and Y. Kim, Classication of heart diseases based on ECG Int. signals using long short-term memory, doi: , ., /EMBC., ., . in Proc. , th Annu.

M. A. Malik. (,). Internet of Things Healthcare Market, Allied , Online, . Available: Market Research. Accessed: Oct. , , .

X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, A hybrid classication system for heart disease diagnosis based on the Art. no. , .

A. L. Bui, T. B. Horwich, and G. C. Fonarow, Epidemiology and risk , .

C. Xu, J. M. Schmitt, S. G. Carlier, and R. Virmani, Characterization of atherosclerosis plaques by measuring both backscattering and attenuation no. , , Art. no. , doi: ,