4th International Conference on Electrical Information and Communication Technology

Ahsanullah University of Science and Technology

Dept. of Computer Science and Engineering

Artificial Prognosis of Cardiac Disease using an NN: A Data-scientific Approach in Outlier Handling

Shithi Maitra, Tonmoy Hossain, Abdullah Al-Sakin and Faisal Muhammad Shah

Outline

- Introduction
- Review of Related Literature
- Proposed Methodology
- Proposed Neural Network
- Experimental Results and Comparison
- Conclusion and Future Plans

Introduction

Introduction: Severity of the Problem

- According to World Health Organization, cardiovascular diseases claim 19.9 million lives per year.
- Greater than 75% of these deaths victimize natives of low and middle-income countries.
- Heart attacks and strokes alone engulf 85% of such deaths.
- According to American Heart Association, most of the cases are addressable if detected early.

Introduction: Challenging Problem Includes 3 steps

- careful examination of medical history
- performing focused physical examinations
- deciding which diagnostic system will provide a complete diagnosis

Review of Related Literature

Review of Related Literature

- Recent scientific literature (Fig. 1) focused on designing artificially intelligent systems to diagnose cardiac infirmity can be reviewed along four paradigms:

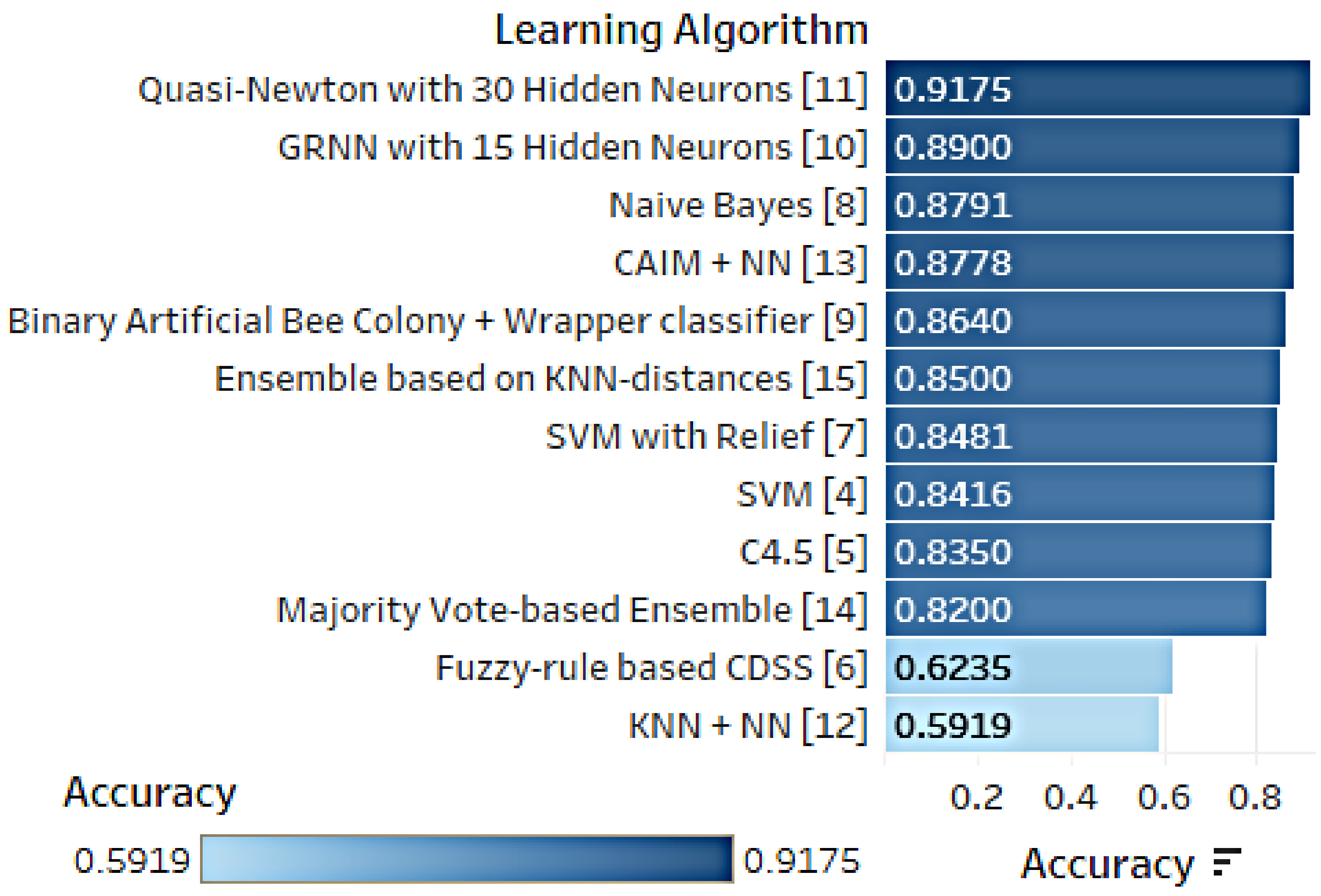

Fig. 1. comparison among related researches

Review of Related Literature

- Nonparametric Machine Learning Algorithms
- Parametric Machine Learning Algorithms
- Neural Network-based Approaches
- Hybrid and Ensemble-based Approaches

Proposed Methodology

Proposed Methodology

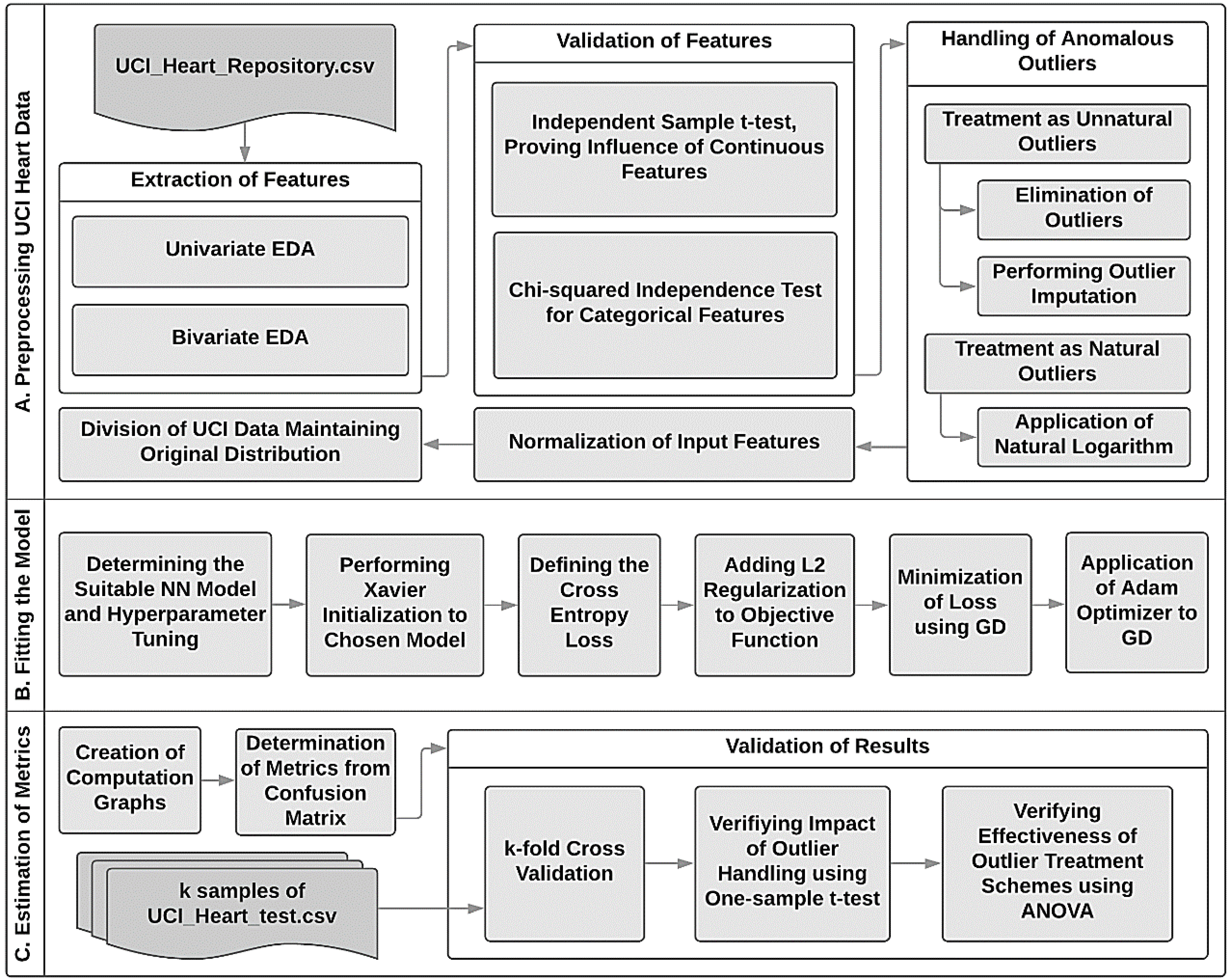

Fig. 2. workflow for the proposed detection of a possible cardiovascular disease

Extraction of Features using Exploratory Data Analysis

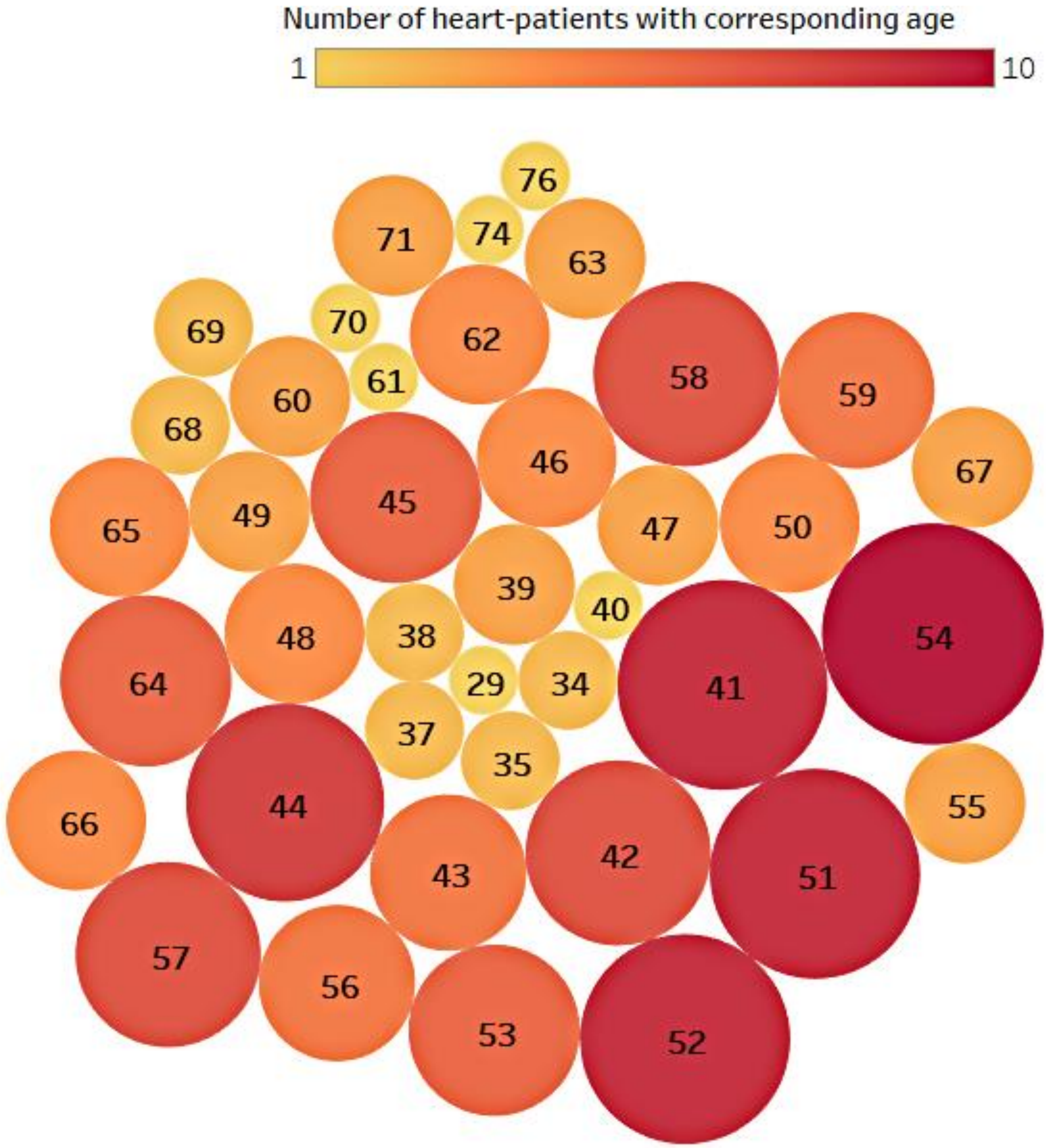

Fig. 3. age groups of heart patients with the size of bubbles representing the number of affected subjects

Extraction of Features using Exploratory Data Analysis

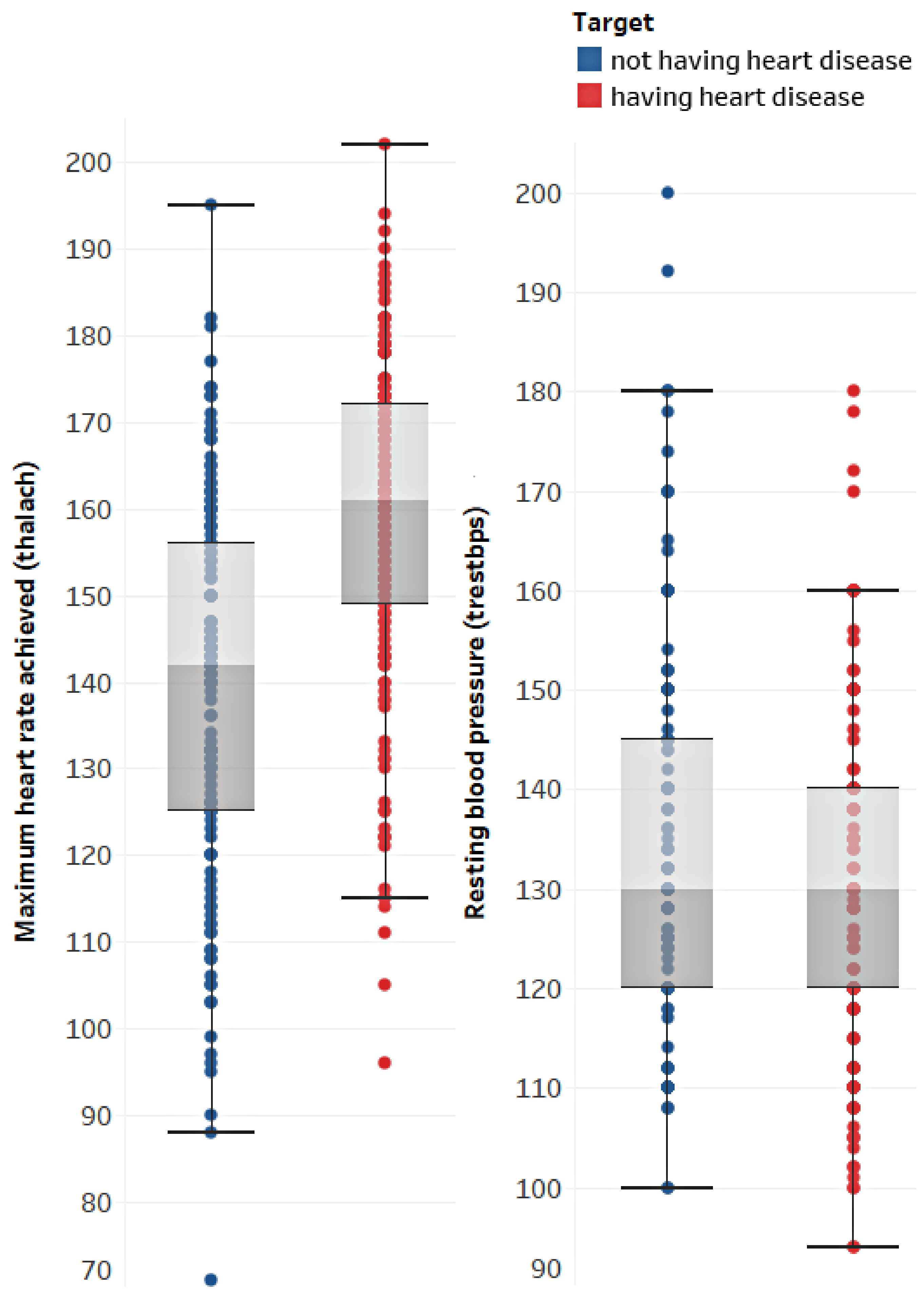

Fig. 4. comparative dispersion of maximum heart rate and resting BP in normal and diseased subjects

Extraction of Features using Exploratory Data Analysis

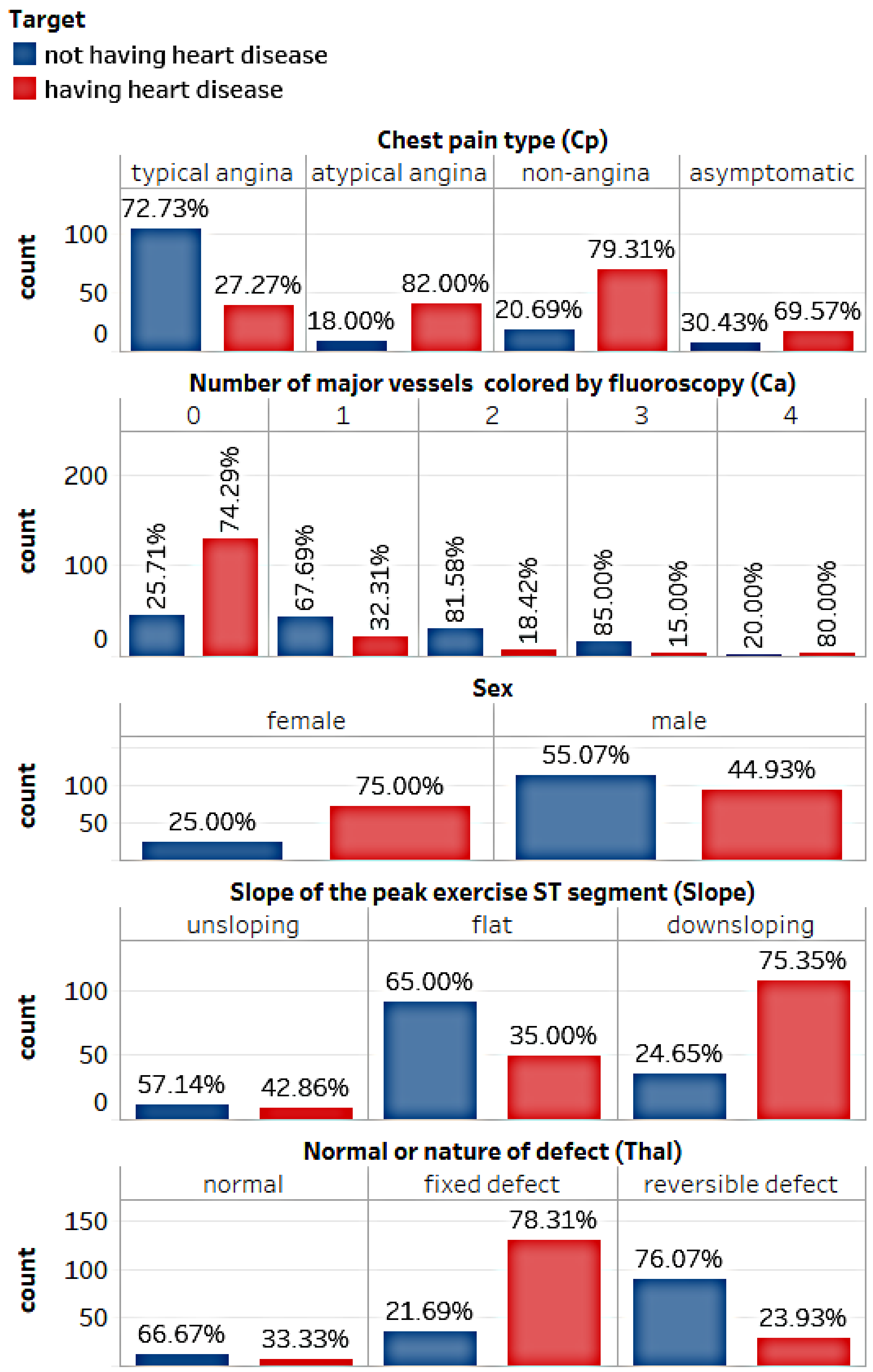

Fig. 5. ratios of subjects portraying each level of chest pain, colored vessel, gender, slope and defect

Extraction of Features using Exploratory Data Analysis

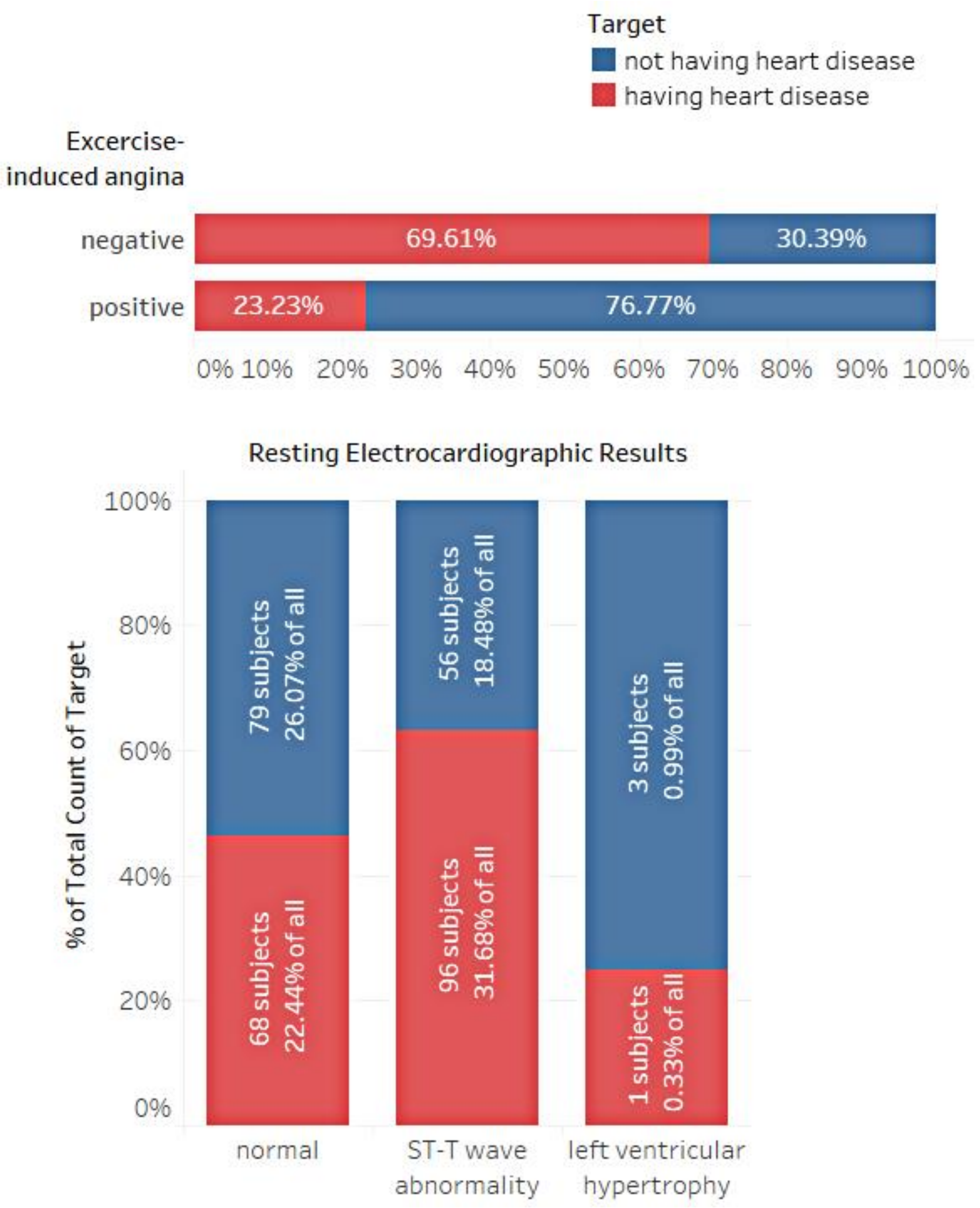

Fig. 6. impact of ECG results and anginal pain among dichotomized subjects

Validation of Extracted Features

TABLE I
FINDINGS OF WELCH TWO SAMPLE T-TEST

continuous features	t-score	degrees of freedom	p-value	H_o: $\mu_1 = \mu_2$	H_a: $\mu_1 \neq \mu_2$
age (years)	4.0797	301	5.78E-05	reject	retain
max heart rate (bpm)	-7.953	269.9	5.02E-14	reject	retain
resting blood pressure (mmHg)	2.5083	272.56	0.01271	reject	retain
oldpeak (ST depression)	7.9386	215.68	1.11E-13	reject	retain
serum cholestoral (mg/dl)	1.4948	298.03	0.136	retain	reject

TABLE II
RESULTS OF PEARSON'S χ^2-TEST

Pearson's χ^2-test					
discrete features	χ^2	degrees of freedom	p-value	H_o : no association	H_a : association exists
chest pain type, target	81.686	3	$< 2.2E-16$	reject	retain
nature of defect, target	85.304	2	$< 2.2E-16$	reject	retain
major vessels colored, target	74.367	4	2.71E-15	reject	retain
slope, target	47.507	2	4.83E-11	reject	retain
resting ECG, target	10.023	2	0.006661	reject	reject
Pearson's χ^2-test with Yates' continuity correction					
discrete features	χ^2	degrees of freedom	p-value	H_o: no association	H_a: association exists
sex, target	22.717	1	1.88E-06	reject	retain
exercise induced angina, target	55.945	1	7.45E-14	reject	reject
fasting blood sugar, target	0.10627	1	7.44E-01	retain	reject

Proposed Neural Network

Proposed Neural Network

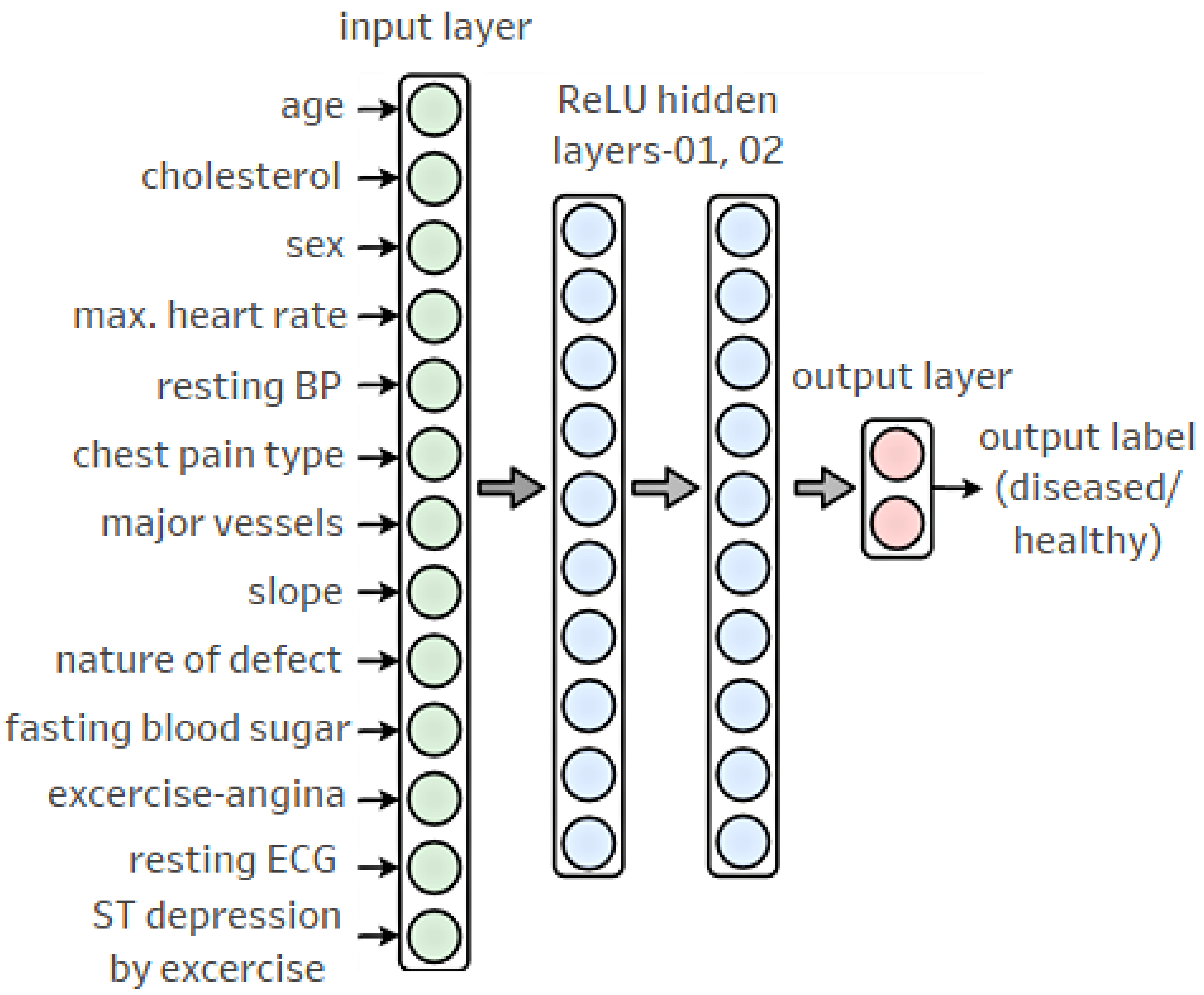

Fig. 7. the proposed 3-layer NN to detect cardiovascular disease

Proposed Neural Network

- **number of layers, neurons**: Having just 303 examples, two hidden layers have been chosen to preempt overfitting with as many as ten hidden neurons, to preclude underfitting.
- **learning rate, α**: A small learning rate of 0.001 was chosen to prevent overshooting across minima.
- **regularization parameter, λ**: Inversely proportional to overfitting, this value has been set to 0.08.
- **number of epochs**: Training through a large 200 epochs quantified the highest refined parameters.
- **size of minibatch**: Minibatches of 64 tuples were used so as to not occupy a great share of the primary memory as done otherwise in batch gradient descent.

Description, Division and Distribution of the Cleveland UCI Dataset

- the dataset contains 138 tuples (45.545%) of non-diseased and 165 tuples (54.455%) of diseased samples
- we make an 80%-20% division of training and cross-validation data
- we assure a fair distribution of the tuples within the sets using R

Treatment as Natural Outliers

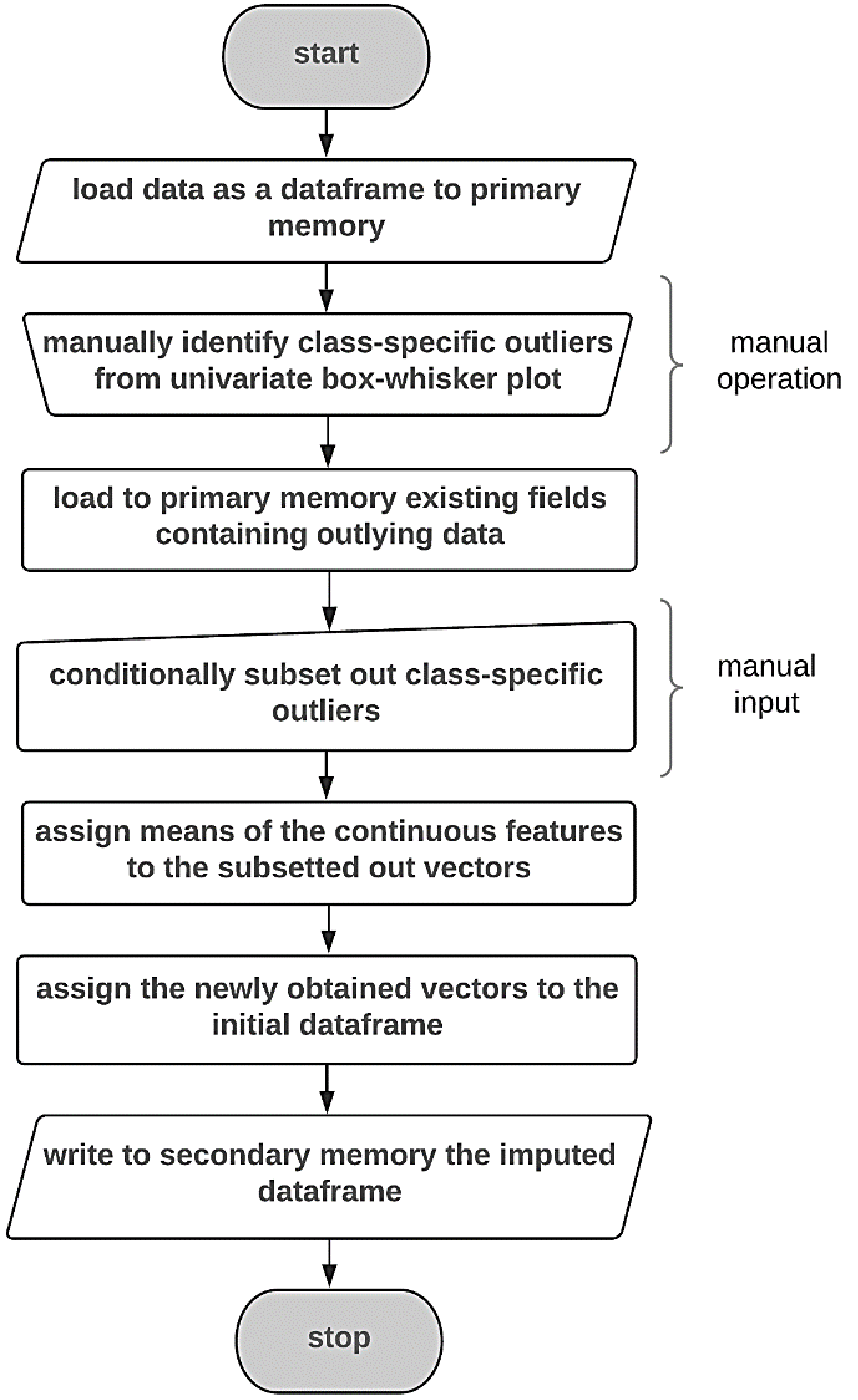

Fig. 8. the algorithm for class-specific mean imputation of natural outliers

Experimental Results and Comparison

Experimental Results

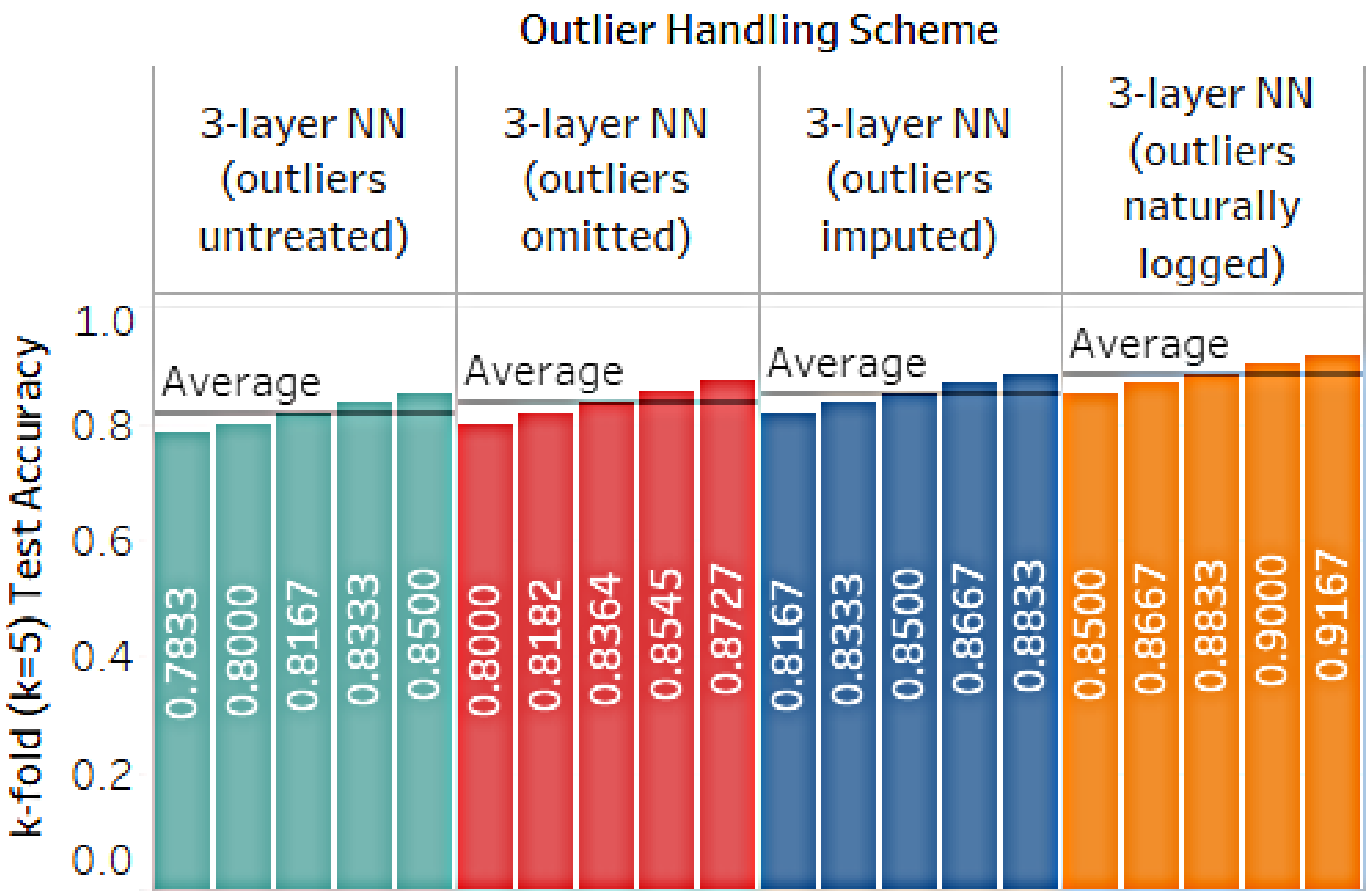

Fig. 9. comparison among proposed schemes' average performance measures

Comparison with Existing Literature

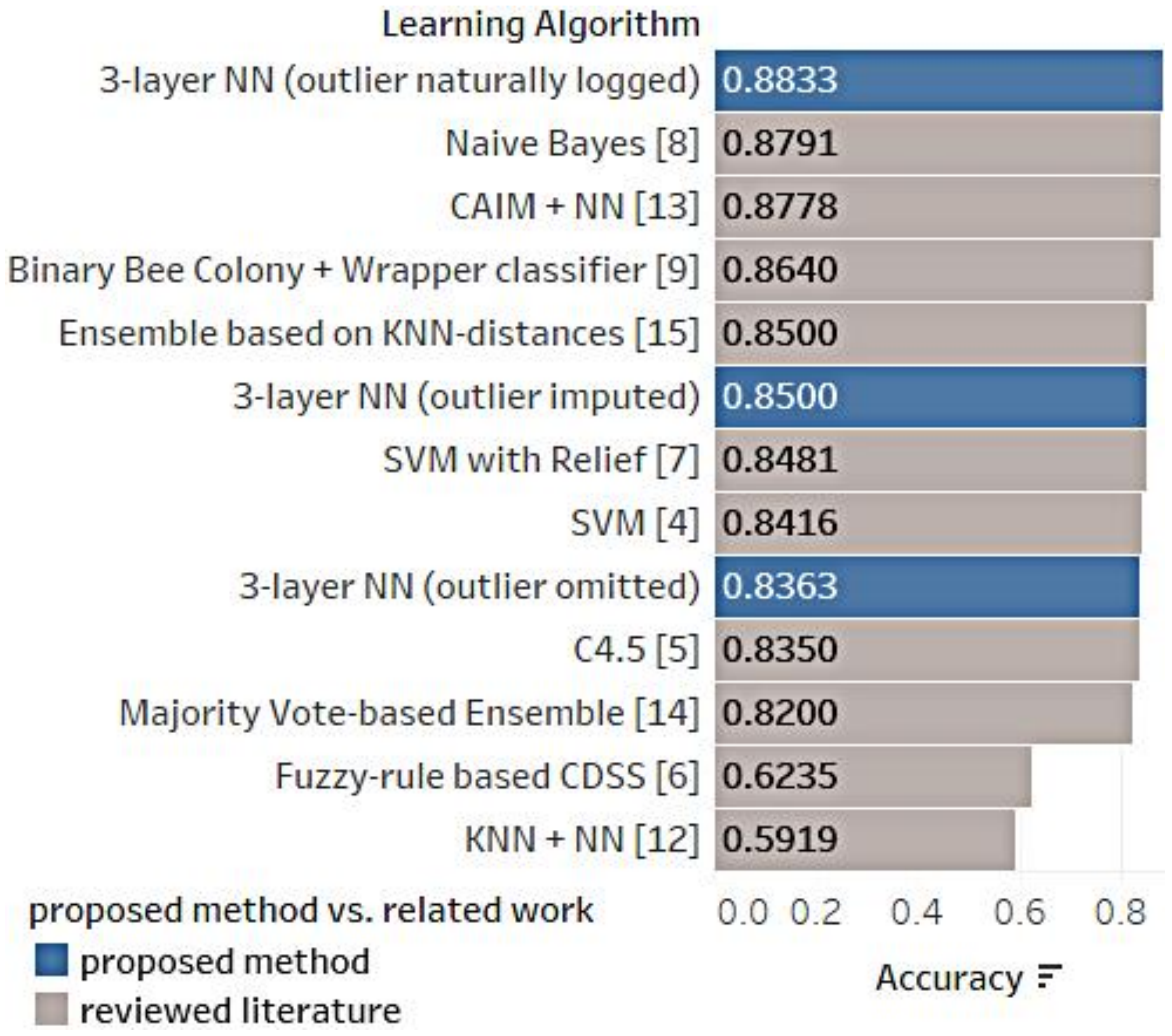

Fig. 10. a comparative analysis between our methodology and reviewed literature

Conclusion and Future Plans

Conclusion

- ✓ The paper endorses exploratorily analytical feature filtering followed by effective engineering of the dataset.
- ✓ The research intensely tested and verified all alterations in the outlier-handling schemes by statistical tests—proving the need for noisy data-management.
- ✓ The study k-fold cross-validated its results, proving statistical consistency on test data, giving an unbiased evaluation.

Future Plans

- ✓ More mathematical mappings can be tried out to handle outliers
- ✓ This approach can be extended to more medical problems
- ✓ The study k-fold cross-validated its results, proving statistical consistency on test data, giving an unbiased evaluation.

Thank You!

References

[1] https://www.who.int/cardiovascular diseases/en/ [Accessed on 20 July, 2019].
[2] Wyatt, Jeremy, and David Spiegelhalter. "Field trials of medical decision-aids: potential problems and solutions." Proceedings of the annual symposium on computer application in medical care. American Medical Informatics Association, 1991.
[3] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases"
[4] Sen SK. Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms. International Journal of Engineering And Computer Science. 2017 Jun;6(6).
[5] Chaki D, Das A, Zaber MI. A comparison of three discrete methods for classification of heart disease data. Bangladesh Journal of Scientific and Industrial Research. 2015 Dec 11;50(4):293-6.
[6] Anooj PK. Clinical decision support system: risk level prediction of heart disease using decision tree fuzzy rules. Int J Res Rev Comput Sci. 2012 Jun;3(3):1659-67.
[7] Takci H. Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences. 2018 Jan 27;26(1):1-0.
[8] Dulhare UN. Prediction system for heart disease using Naive Bayes and particle swarm optimization. Biomedical Research. 2018;29(12):2646-9.

References

[9] Subanya B, Rajalaxmi RR. Artificial bee colony based feature selection for effective cardiovascular disease diagnosis. International Journal of Scientific & Engineering Research. 2014;5(5):606-12.
[10] Khanna D, Sahu R, Baths V, Deshpande B. Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease. International Journal of Machine Learning and Computing. 2015 Oct 1;5(5):414.
[11] Karim H, Niakan SR, Safdari R. Comparison of Neural Network Training Algorithms for Classification of Heart Diseases.
[12] Malav A, Kadam K. A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means. International Journal of Pure and Applied Mathematics. 2018;118(8):103-10.
[13] INAN O, CIFCI ME. Diagnosing Diseases via Artificial Intelligence Powered by New Hybrid System. International Journal of Engineering Science. 2018 Jun;18534.
[14] Bashir S, Qamar U, Javed MY. An ensemble based decision support framework for intelligent heart disease diagnosis. International Conference on Information Society (i-Society 2014) 2014 Nov 10 (pp. 259264). IEEE.
[15] Pawlovsky AP. An ensemble based on distances for a kNN method for heart disease diagnosis. 2018 International Conference on Electronics, Information, and Communication (ICEIC) 2018 Jan 24 (pp. 1-4). IEEE.
[16] Ng, Andrew. "CS229 Lecture notes." CS229 Lecture notes 1 (2000) [Accessed on 29 November, 2018].