J. Cardiovasc Disease Res., 2020; 11(4): 26 – 31
A Multifaceted Peer Reviewed Journal in the field of Cardiology
www.jcdronline.org | www.journalonweb.com/jcdr
P – ISSN: 0975 – 3583 Online: 0976 – 2833

Cardio-Vascular Disease Classification Using Stacked Segmentation Model and Convolutional Neural Networks

[1]G. Charlyn Pushpa Latha, [1]S. Sridhar, [2]S. Prithi, [3*]T. Anitha
[1]Saveetha Institute of Medical &Technical Sciences (Deemed to be University), Chennai, Tamil Nadu, India.
[2]Rajalakshmi Engineering College (Autonomous), Chennai, Tamil Nadu, India.
[3]Makeit Technologies (Center for Industrial Research), Coimbatore, Tamil Nadu, India.*
Avinashilingam Institute for Home Science & Higher Education for Women (Deemed to be University), Coimbatore, Tamil Nadu, India.
Corresponding author: T. Anitha

ABSTRACT

Various decision support system based on Artificial Neural Networks have been extensively used for predicting cardio-vascular disease. But, some of the investigations concentrate on pre-processing the features. Here, this work focuses on feature refinement with segmentation and removal of problems related to prediction model. The problems are related to over-fitting and under-fitting. By avoiding these problems, the proposed model shows superior functionality while considering the available dataset. For eliminating the un-necessary parts of input data, de-noising stacked encoder is used for configuring the Convolutional Neural Networks with testing and training data. The anticipated model is compared with existing approaches and reported that this model outperforms the existing approaches for predicting heart disease. The anticipated model acquires finest prediction accuracy. The results are seems to be more promising while compared to the other. The findings based on this study recommend that this diagnostic system is utilized for physicians to predict heart disease accurately. Simulation is done in MATLAB environment.

Keywords: Cardio-vascular Disease, CNN, Feature Refinement, Prediction Accuracy, Segmentation.

Correspondence:
T. Anitha
Makeit Technologies (Center for Industrial Research), Coimbatore, Tamil Nadu, India.
Avinashilingam Institute for Home Science & Higher Education for Women (Deemed to be University), Coimbatore, Tamil Nadu, India.

Submitted: 29-07-2020
Revision: 15-08-2020
Accepted Date: 01-09-2020

DOI: 10.31838/jcdr.2020.11.04.05

INTRODUCTION

Death that happens all over the world in past few decades shows the primary cause due to heart disease [1]. It is measured as a common health crisis, and it essential cause is that the sum of blood needed to fulfill body requirements is not pumped well by the heart [2]. It is due to unhealthy food condition, genetic issues, mental stress and lack of exercise. It influences the person slowly and it is extremely complex to handle until it is progressed to some dangerous level [3]. If the essential remedial measures are not considered, it may adversely influence the subject. Therefore, heartdisease prediction is essential; therefore the patient may take care of their body condition when heart disease is sensed [4]. The most general way to acquire all the data associated with patient is to utilize information system to handle the patient's health in precise manner.

Heart disease is significantly due to certain risky factors classified as non-changeable factors and changeable factors [5]. The things that come under changeable forms are cholesterol level, blood pressure and so on while non-changeable factors may include non-changeable factors like patients' sex, age and history. There is some substantial amount of data that is not being utilized adequately known for support clinical decisions [6]. This widen information is also utilized to identify heart disease more appropriately in earlier prediction stages. Even though, heart disease diagnosis is performed by invasive cardiology based approaches such as angiography. It is based on patients' medical history and higher level of technical expertise to examine related symptoms [7]. As well, it is extremely a difficult process with superior costs, therefore non-invasive approaches such as decision support systems dependent on deep learning and machine learning has to be utilized to get rid of this crisis [8].

Various investigators have determined diverse techniques and design methodologies for expert system that are competent of identifying heart disease in patients' earlier. There are some significant enhancements in quality of these anticipated diagnostic approaches. This motivates to model a de-noising segmentation model with auto-encoders that has an ability to deal with all the under-fitting and over-fitting problems and optimization with network configuration crisis. Therefore, heart disease diagnosis can be enhanced.

In the model [9], the ultimate objective is to resolve the crisis that is considered to be crucial cause for performance degradation with diagnostic system. This may prevents the prediction accuracy for attaining superior level that is needed. One factor is that how to handle data, where it avoids under-fitting and over-fitting. Another problem is how to determine the optimal network configuration with auto-encoder. Therefore, optimization problem can be eliminated. When the model performs effectually with training data, it leads to over-fitting as it picks lesser amount of information from training data and while applying to testing data that outcomes are more appropriate.

Subsequently, when the model is not learned more effectually for training data, it leads to under-fitting as an outcome both training and testing data demonstrates poor outcomes. When the network configuration is not determined properly and shows some inappropriate features they may not contribute more appropriately as a reason behind these factors. Due to this crisis, computational complexity will be raised and prediction time for heart disease may raises. To eliminate features that are redundant or noisy, this work considers stacking based auto-encoder as it is more effectual for determining optimization features. The subsequent step is to anticipate a network model for segmentation and to optimize network configuration. This work uses stacking based auto-encoder which is competent of network tuning to find optimal solution and to enhance system performance.

In the final process, experimentations are carried out and various evaluation metrics are considered that includes precision, accuracy, sensitivity and specificity and F1 score. The outcomes were validated with cross validation process. To perform testing and training, UCI machine learning repository for heart disease dataset are utilized. This dataset is chosen due to its most extensive investigation towards heart disease prediction for showing the efficiency of the model. The experimentations are carried out in MATLAB over a PC with Intel Core processor with 2.90 GHz CPU.

RELATED WORKS

Before the existence of CNN, most of the image processing functions like image segmentation, classification and object detection are based on handcrafted features like local binary patterns, Haar functions and with some complex classification functions like (cascaded classifiers, SVM and RF). Deep learning offers end-to-end outcomes for all these tasks as it performs both classification and feature selection concurrently. Image segmentation is considered as an essential part that is in the pipeline order.

Based on up-sampling function, existence of deep learning approach for segmenting images are partitioned into two factors. Initially, this is may not apply up-sampling as it may uses patches as input and categorize the central pixels. This process is performed simultaneously for all pixels. It works slowly as each single patch needs independent feed-forward computation. Moreover, this problem is resolved partially using weighted function.

Some other prevailing models may use various approaches like deep learning and machine learning for prediction. Various approaches comprises of SVM, ANN, Fuzzy concept, decision tree and DNN and LSTM for recognizing the symptoms of heart disease. An appropriate decision making enhances significance of expert systems and it may slightly reduce death rate due to its diagnostic systems. Multi-modal prediction based model is provided for some un-structural and structural model and these models may work effectually for recognizing chronic disease. The integration of FL and NN was anticipated in [10] that offers effectual contribution score for all attributes in recognizing heart disease. The outcomes show enhanced accuracy in contrary to CNN. It utilizes essential statistics of some patients' medical records. Authors [11], used three finest model to enhance overall prediction accuracy. They are Naïve Bayes, AdaBoost and NB and the investigators determined that this model works effectually in recognizing heart disease with superior accuracy than traditional models. Author in [12], anticipated fuzzy diagnostic approach using five diverse factors to acquire prediction accuracy. Ruzzo-Tompa is merged with NN to reduce data dimensionality and classification with neocognition NN. Some research concentrates on over-fitting crisis for addressing NN based training for all feature ranking and it is extended to a competency of heart disease diagnosis. The integration of adaptive neuro-fuzzy interference system and multi kernel model is used for disease prediction. The two-fold model is used for parameter separation among healthy and non-healthy patients. The outcomes are provided to ANFIS for classification purpose. Some other effectual and expert diagnostic model for heart disease prediction is given in [13], with two SVM models. It is utilized for eliminating irrelevant features from feature set and others are used for prediction. For model optimization, hybridization approach is used for grid search, therefore both models has to be simultaneously optimized.

Some investigators anticipated an ensemble NN model for predicting heart disease and to acquire 90% accuracy, 80% sensitivity and 95% specificity. Author in [14], modeled a novel hybridization system for Fuzzy and ANN analytic process for disease prediction. Fuzzy and ANN based system may acquire prediction accuracy. This may check feasibility of ANFIS [15] for heart disease prediction and acquired classification accuracy superiorly.

METHODOLOGY

This section discusses about the segmentation process carried out for predicting noise over dataset images. Here, stacked auto-encoder is used for segmentation and the dataset has been partitioned into testing and training for validation purpose.

a. Dataset Description

The heart disease dataset is utilized to perform the experimentation which is publicly available in UCI-machine learning repository. This dataset comprises of roughly about 300 instances with missing values. Various features are considered from the heart disease dataset with diverse values with a range of 0-1 where training can be interpreted more effectually. Generally, dataset has been partitioned into three diverse sub-sets termed as testing, training and validation. To train these datasets, training subset is utilized. To optimize classifier parameters, subset validation and test validation is used.

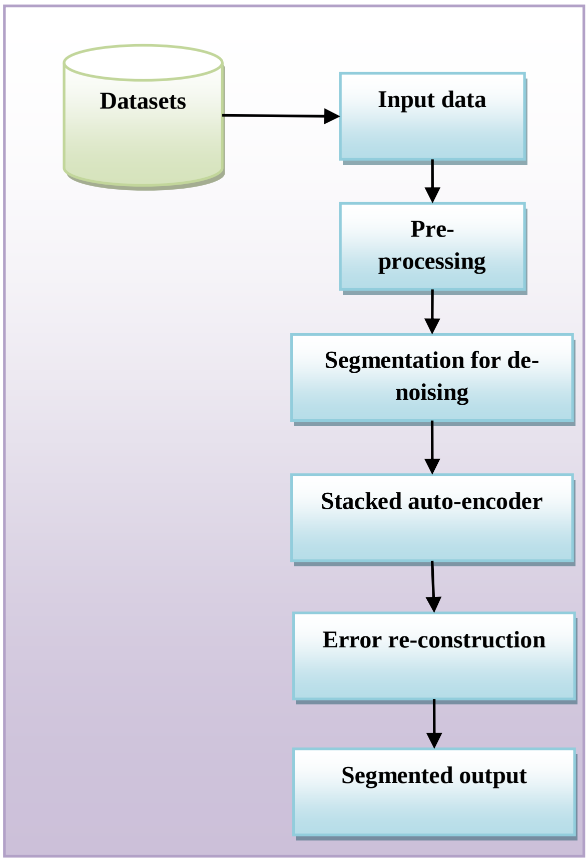

Fig. 1: **Flow diagram of proposed model**

Here, deep learning approaches are extensively utilized for performing segmentation tasks automatically. Specifically, heart region from MRI images are utilized in CNN where data acquired from healthcare dataset. The contours are eliminated from cardiac segmentation process which is needed during analysis process. In deep learning process, various approaches are used for examining number of available data like analysis, fusion with clinical support system for validation. Initially, the model is formulated for data analysis. These approaches may provide various techniques for probable manipulation of risky factors by re-construction of training strategy. The dataset representation in matrix format is expressed as below:

$$X = [x_1, \ldots, x_m] = \begin{bmatrix} x_{1,1} & \ldots. & x_{M,1} \\ \ldots. & \ldots.. & \ldots. \\ x_{1,N} & \ldots. & x_{M,N} \end{bmatrix} \quad (1)$$

Here, mapping function is given as $f: X \rightarrow Y$ is followed by various health care records to allocate individual data with appropriate patients' group. The risk prediction for cardio-vascular disease is based on diverse segmented approach. Asymmetric NN is utilized for data fusion and analysis. The anticipated stacked auto-encoder is utilized for analyzing dataset function. More specifically, the learning process commences from training set with auto-encoder for analyzing the flow of blood with de-noised and noised heart image. The input dataset is considered as a binary version of X to attain hidden layer are represented in default. There are some output nodes that is equal to input number and sample vector length of individual patient's for analysis. Non-linear function of input and vector functionality is determined by hidden vector is specified for encoding and it is specified as in Eq. (2):

$$e = g(V^{h-1}y + a^h) \quad (2)$$

The encoder parameters are specified as $\rho^h = (V^h, a^h)$. When decoding, hidden vector is mapped to re-construct input vector by non-linear activation function is specified as in Eq. (3):

$$X' = d(V^t e + a^t) \quad (3)$$

Decoder function is specified as $\rho^t = (V^t, a^t)$.With all patients' sample, y_j is equal to hidden vector for sampling and reconstructed as X to size $\tilde{y}_j$. By reducing reconstruction error, parameter optimization is determined as in Eq. (4):

$$G(\rho) = \frac{1}{M}\sum_{j=1}^{M} K(y_j, y_j') \quad (4)$$

$$G(\rho) = \frac{1}{M}\sum_{j=1}^{M} K\ (y_j, e_\rho[g_\rho(y_j')) \quad (5)$$

Here, loss function is specified as $\rho^t = (V^t, a^t)$.For all individual samples, y_j is sample to hidden vector and reconstructed to X to size $\tilde{y}_j$ sample. By diminishing re-construction error, parameters are optimized as Eq. (6):

$$K(x_j, x_j') = \sum_{i=1}^{N} (x_{ij} \log x_{ji}' + (1 - x_{ij}) \log(1 - x_{ji}') \quad (6)$$

To reduce cost with Eq. (6), parameter vector is initiated for zero and optimization process is used with smaller values.

Algorithm 1:

1. Perform pre-processing with stacked auto-encoder
2. Initialize input variables x, y
3. Compute total input data used for validation with individual matrix and image de-noising
4. Compute encoding process with Eq. (2)
5. Revise hidden vector parameters
6. Compute non-linear activation function
7. Model a decoder function with sample functions
8. Revise values for eliminating re-construction error
9. Design the model by eliminating reconstruction error
10. End for
11. End process

Measuring risk level

1. Determine risk level with two samples.
2. Set sample values
3. Compute risk function
4. Evaluate inter- and intra-segmentation
5. Update objective function
6. Evaluated outcomes for inter and intra segmentation
7. Perform normalization for evaluating risk level
8. Identify risk level
9. End for
10. End process

EVALUATION METRICS

To compute the performance of anticipated prediction approach, diverse evaluation metrics like specificity, accuracy, sensitivity and MCC is utilized. Accuracy is depicted as percentage of appropriately classified individuals in test dataset. Sensitivity is depicted as information regarding percentage of being appropriately classified patients while specificity determines information regarding appropriately classified healthy individual. The Equation of these metrics is given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where TP is number of true positives, FP is number of false positives, TN is number of true negatives, FP is number of false negatives. MCC is utilized for performing statistical analysis using binary classification. It is for measuring the test accuracy. It returns value among -1 and 1 where '1' specifies prediction and '-1' specifies worst predictions. Here, segmentation is done with stacked auto encoder for handling noise over images. Here, 207 samples are used for training and 90 samples are used for testing. Testing samples are used to train CNN model. To validate the performance of anticipated method enhances performance traditional CNN. This simulated CNN model uses features set by optimizing search approach. The experimental outcomes show that finest accuracy 90% can be attained using CNN on full features. The network configuration comprises hidden layers and neurons in hidden layer. From experimental outcomes, it is clearly proven with de-noising noise enhances traditional CNN performance.

Table I: Dataset description

No	Feature	Abbreviation	Type	Description
1	F_1	Age	Continuous	Patients' age
2	F_2	Sex	Discrete	Patients' gender
3	F_3	CPT	Continuous	Chest pain type
4	F_4	RBP	Continuous	Resting blood pressure
5	F_5	SCH	Continuous	Serum cholesterol
6	F_6	FBS	Discrete	Fasting blood sugar
7	F_7	RES	Discrete	Resting Electro-cardio graphic
8	F_8	MHR	Continuous	Maximal heart rate
9	F_9	EIA	Discrete	Exercise induced angina
10	F_{10}	OPK	Continuous	ST depression
11	F_{11}	PES	Discrete	Peak slope ST segment
12	F_{12}	VCA	Continuous	Number of major vessels
13	F_{13}	THA	Discrete	Thalassemia

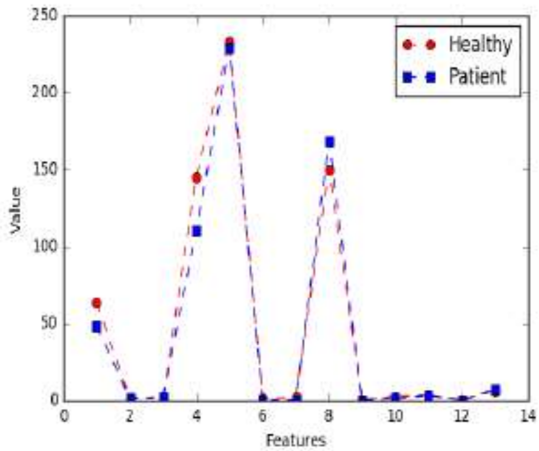

Fig. 2: Validating healthy and normal patients

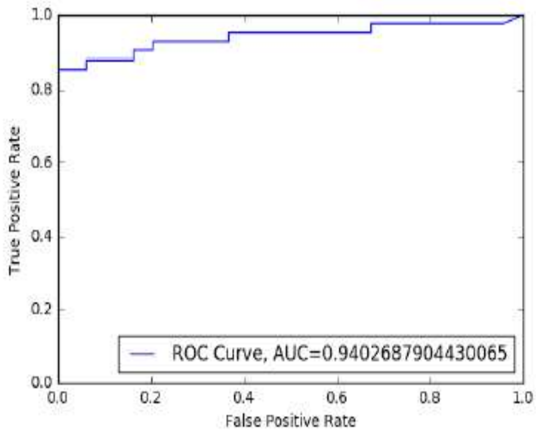

Fig. 3: ROC computation

To validate efficiency of anticipated hybrid model, ROC and AUC are used. It is utilized to analyze classifier output quality. It is for plotting true positive rate and true negative rate. It validate the enhancement of performance of anticipated model, comparative study is performed with well known machine learning models with traditional models. To evaluate specificity and sensitivity, classifier is used for dataset with graphical structure. CNN with network optimization is used to test performance of de-noising. The performance by eliminating features has noisy characteristics that are not contributed to enhance heart disease prediction. Experimental outcomes show superior accuracy and finest evaluation metrics than models.

Table II: Performance metrics

Epochs	1	8	15	20
Accuracy	67	75	80	87
Precision	69	78	83	89
Recall	90	85	75	70
F-measure	66	72	82	91

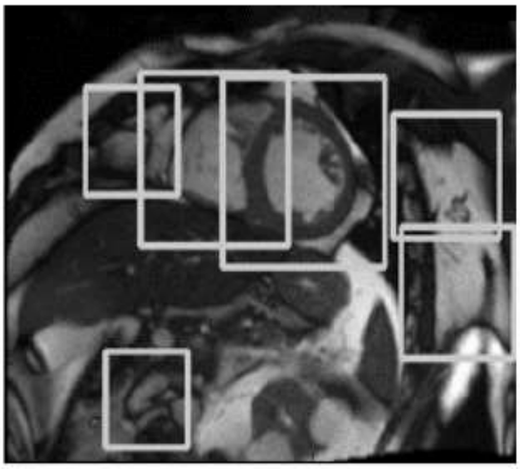

Fig. 4: Segmented region

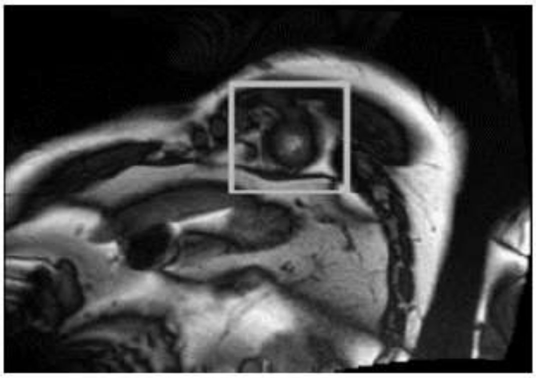

Fig. 5: Segmenting noisy region with stacker

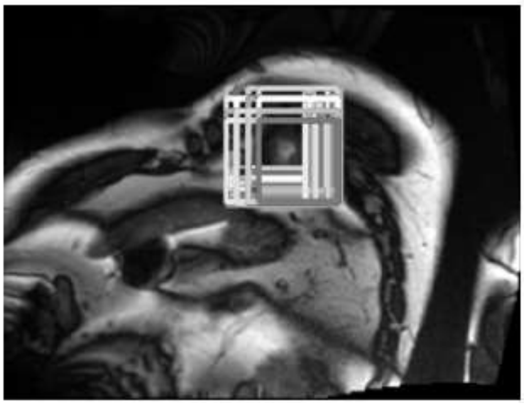

Fig. 6: Prediction of noisy region

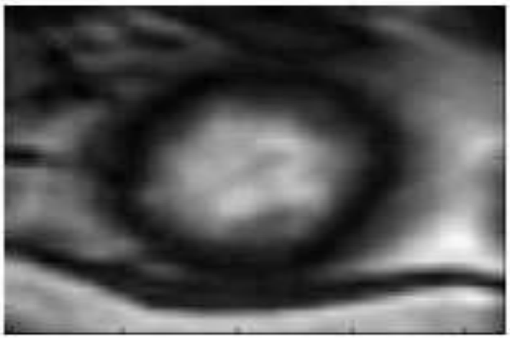

Fig. 7: Noisy region

Table I depicts the dataset descriptions with various parameters like age, gender, sex, CPT, RBP, SCH, FBS, RES, MHR, EIA, OPK, PES, VCA, THA. These thirteen features are considered to more essential for predicting cardio-vascular disease. The values associated with this are either continuous or discrete. These values are determined based on accuracy validation. Table II depicts the performance metrics used for the proposed model for various epochs like 1, 8, 15 and 20. The accuracy is 67%, 75%, 80%, 87% respectively. Precision is 69%, 78%, 83%, 89% respectively. Recall values are 90%, 85%, 75% and 70% respectively. Similarly, F-measure values are 66%, 72%, 82% and 91% respectively.

Fig 2 specifies the normal patient flow with the abnormal patient flow. This flow is based on the continuous and discrete values generated and maintained over the dataset already. Fig 3 depicts the ROC and AUC measure of proposed model. The AUC value is approximately 0.94. Fig 4 and Fig 5 shows the segmented regions of proposed model with stackers. The noises over the input images are eliminated with this stacker model. The segmented region shows the noise occurrence from the sample input data. The noisy region is identified and further prediction process is carried out. The anticipated model is considered to be more effectual than the conventional approaches.

CONCLUSION

Here, this work models an automated diagnostic model for predicting heart disease. The anticipated model uses stacked auto-encoder for segmenting images to eliminate noise. The strength of anticipated model was computed with various metrics like specificity, sensitivity, accuracy, ROC, AUC and MCC. The performance of anticipated approach is compared with conventional models. The experimental analysis concluded that decision making quality is improved when predicting heart disease. The anticipated model acquires superior detection accuracy for disease prediction however the present study has not analyzed time complexity of this segmentation process. In future, time complexity has been given higher concentration and it is also considered as an essential factor for clinical applications. Some other limitations are based on optimal width of hidden layers using the proposed model. Thus, faster algorithms will be used in future.

CONFLICT OF INTEREST

None

REFERENCES

1. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer methods and programs in biomedicine* 2017; 141: 19-26.
2. Olaniyi EO, Oyedotun OK, Khashman A. Heart diseases diagnosisusing neural networks arbitration. *Int. J. Intell. Syst. Appl.,* 2015; 7(12): 75_82.
3. Manogaran G, Varatharajan R, Priyan MK. Hybrid recommendation system for heart disease diagnosis based on multiple kernel learningwith adaptive neuro-fuzzy inference system. *Multimedia Tools Appl.,* 2018; 77(4): 4379_4399.
4. Acharya UR, Fujita H, Lih OS, Adam M, Tan JH, Chua CK. Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network. *Knowl.-Based Syst.,* 2017; 132: 62_71.
5. Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G. An integrated decision support system based on ANN and fuzzy_AHPfor heart failure risk prediction. *Expert Syst. Appl.,* 2017; 68: 163_172.
6. Paul AK, Shill PC, Rabin MRI, Murase K. Adaptive weightedfuzzy rule-based system for the risk level assessment of heart disease. *Appl. Intell.,* 2017; 48(7): 1739_1756.
7. Pourtaheri ZK, Zahiri SH. Ensemble classi_ers with improved over_tting. *In Proc. IEEE 1st Conf. Swarm Intell. Evol. Comput. (CSIEC)* 2016.
8. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models withelectronic health records data: A systematic review. *J. Amer. Med. Inform. Assoc.,* 2017; 24(1): 198-208.
9. Huang Z, Dong W. Adversarial MACE prediction after acute coronary syndrome using electronic health records. *IEEE J. Biomed. Health in format* 2019; 23(5): 2117-2126.
10. Ng K, Stewart WF, De Filippi C, Dey S, Byrd RJ, Steinhubl SR, Daar Z, Law H, Pressmnan AR, Hu J. Data-driven modelingof electronic health record data to predict prediagnostic heart failurein primary care. *J. Patient-Centered Res. Rev.,* 2016; 3(3): 200.
11. Hernesniemi SM, Tynkkynen JA, Lyytikäinen LP, Mishra PP, Lehtimäki T, Eskola M, Nikus K, Antila K, Oksala N. Extensive phenotype data and machine learning in prediction of mortalityin acute coronary syndrome_The MADDEC study. *Ann. Med.,* 2019; 51(2): 156-163.
12. Huang TMC, Dong W. MACE prediction of acute coronarysyndrome via boosted resampling classification using electronic medicalrecords. *J. Biomed. Inf.,* 2017; 66: 161_170.
13. Digumarthy SR, De Man R, Canellas R, Otrakji A, Wang G, Kalra MK. Multifactorial analysis of mortality in screening detected lung cancer. Journal of Oncology 2018: 7.
14. Lessmann N, van Ginneken B, Zreik M, De Jong PA, De Vos BD, Viergever MA, Isgum I. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Transactions on Medical Imaging* 2018; 37(2): 615–625.
15. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016: 770-778.
16. Rajendran T, Sridhar KP. An Overview of EEG Seizure Detection Units and Identifying their Complexity- A Review. *Curr Signal Transd Ther*., 2019. http://doi.org/10.2174/1574362413666181030103616.
17. Rajendran T, Sridhar KP. Epileptic seizure classification using feed forward neural network based on parametric features. *Int J Pharma Res.*, 2018; 10(4): 189-196.
18. Rajendran T, Sridhar KP. Epileptic Seizure-Classification using Probabilistic Neural Network based on Parametric Features. *J Int Pharma Res.*, 2019; 46(1): 209-216.

Cite this article: T. Anitha Cardio-Vascular Disease Classification Using Stacked Segmentation Model and Convolutional Neural Networks. J. Cardiovascular Disease Res., 2020; 11 (4): 26 – 31