Original Article

Diagnosis Of Heart Disease Using Machine Learning Classification Technique In E-Healthcare

[1]Dr. Santosh Kumar , [2]Dr.Satyajee Srivastava , [3]Dr. Shweta Mongia , [4]M. AMSA , [5]Anbumani A , [6]Dr. Amit Chauhan , [7]Ms. Saniya Bhalerao

[1]Associate Professor, Department of Computer Science, Lucknow Public College of Professional Studies, Vinamra Khand, Gomti Nagar, Lucknow -226010, Uttar Pradesh. sanb2lpcps@gmail.com

[2]Professor, School of Computing,University of Engineering and Technology, Roorkee- -247667 Uttarakhand, India. drsatyajee@gmail.com

[3]Associate Professor, Department of computer science and engineering, Manav Rachna international institute of research and studies, Faridabad, India. Shweta.mongia@yahoo.com

[4]Assistant Professor, Department of Artificial Intelligence and Data Science,
M Kumarasamy college of Engineering, Nallagoundam palayam, Namakkal, Pin code-637213,Tamilnadu,India, amsam.ai@mkce.ac.in

[5]Assistant Professor, Department of Information Technology, Kongunadu college of engineering and technology, trichy. anbuinfo2006@gmail.com

[6]Assistant professor, Department of Life sciences, CHRIST (Deemed to be university), Bangalore, Karnataka, India 560029.amit_chauhan777@yahoo.in

[7]Assistant Professor, Department of Pharmaceutics, MET Institute of Pharmacy, Bhujbal Knowledge City, Adgoan,Nashik 422003.Saniya.bhalerao@gmail.com

DOI: 10.47750/pnr.2023.14.S02.80

Abstract

Machine learning is a new discipline in computer science that has given learning a new dimension. It emulates human intelligence in the system by learning from previous data. Due to its self-adaptive capabilities and mathematical foundation, machine learning has several applications in healthcare, sentiment analysis, recommendation systems, natural language processing, information retrieval, gameplay, market research, text recognition, and computer vision. Machine learning may be used to enhance healthcare services. Machine learning decision support systems may be used to diagnose various illnesses. This article's study focuses on detecting heart disease using a machine learning technique. CAD (coronary artery disease) is a kind of heart disease that, if not treated promptly, may result in cardiac arrest. It is the leading cause of death globally.CAD is caused by a blockage in the arteries that provide blood to the heart muscles. Plaque deposition, composed of cholesterol and calcium, causes artery blockages. Atherosclerosis is the medical term for this condition. The endothelium is a thin strip of cells that causes smooth blood flow through arteries. Plaque accumulates in the arterial wall when the endothelium deteriorates. This plaque formation affects the blood flow to the heart, causing blood cells to weaken. CAD worsens with time, and if not treated promptly, symptoms might deteriorate. Plaque accumulates over time, hardening plaque. Platelets may cling together to create blood clots in arteries if the plaque ruptures. It has the potential to entirely cut off the blood flow, resulting in a heart attack. Angiography is a common approach for diagnosing CAD. This procedure is invasive and has certain adverse effects. As a result, non-invasive approaches for diagnosing CAD based on clinical data are required. This article provides an ET-SVMRBF (extra tree support vector machine radial basis function) technique for diagnosing CAD using clinical deed. This approach was validated using the Z-Alizadeh Sani CAD dataset from UCI

(University of California, Irvine). The primary goal for introducing this technology is to minimize mortality by detecting CAD early.

Keywords: Machine Learning, IoT (Internet of Things), Patient Monitoring System, Heart Disease Detection and Prediction

I. Introduction

Heart disease is a leading cause of mortality all over the globe. This illness is responsible for around 31% of all worldwide fatalities. According to the World Health Organization, this illness kills around 17.9 million people yearly. According to the American Heart Association, this condition impacted 121.5 million American adults in 2016 [1-3]. Heart disease should be detected early since it may lead to life-threatening complications. Invasive and non-invasive procedures may be used to diagnose heart disease. Non-invasive procedures include electrocardiograms, phonocardiograms, dynamic electrocardiograms, and echocardiography. Non-invasive techniques of illness diagnosis are less costly and have no adverse effects. However, these strategies are not entirely accurate [4]. As a result, physicians often suggest intrusive procedures.

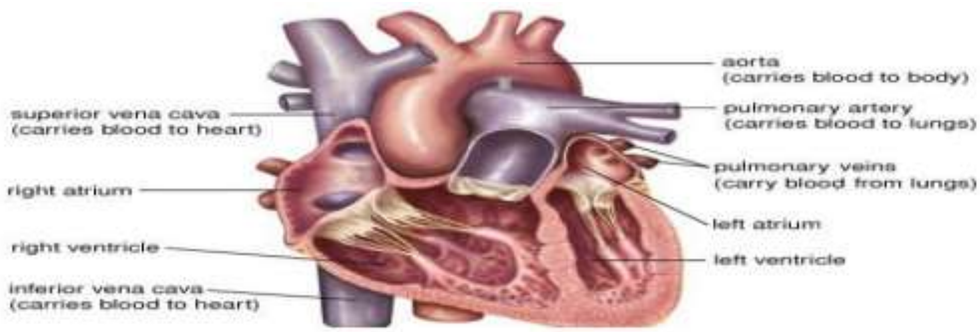

Figure 1: Anatomy of the heart

The heart is responsible for converting deoxygenated blood into oxygenated blood. Arteries are blood arteries that transport blood from the heart to the rest of the body. Veins are blood veins that transport oxygenated blood from the body to the heart. The superior vena cava transports blood from the chest, head, and arms to the right atrium. The inferior vena cava transports blood from the pelvic area, abdomen, and legs to the right atrium. The tricuspid valve transports blood from the right atrium to the right ventricle. The pulmonary artery transports blood from the right ventricle to the lungs. The blood is oxygenated in the lungs. The left atrium gets blood from the lungs and pumps it into the left ventricle through the pulmonary veins.

When arteries get clogged, the heart becomes weak and unable to pump blood via the circulatory system. It is referred to as cardiac failure. The guy dies immediately due to heart failure. Acute or chronic heart failure may occur. Acute cardiac failure strikes unexpectedly. Chronic heart failure is a condition in which the heart cannot pump enough blood. It has an impact on the patient's quality of life. A single medical test cannot detect heart failure. Heart failure is caused by coronary artery disease, cardiac attack, valve disease, hypertension, and other heart disorders. Heart failure may be diastolic, systolic, or left-sided [14]. Left-sided heart failure occurs when the heart cannot provide oxygenated blood to the body.

The left ventricle transports blood from the heart to the rest of the body. If the left ventricle cannot adequately pump blood, the body cannot obtain oxygenated blood, and blood is returned to the lungs instead of being fed to the body. It results in left-sided cardiac failure. The right ventricle pumps blood into the lungs. The lungs convert deoxygenated blood to oxygenated blood. Right-sided heart failure occurs when the heart cannot pump blood to the lungs. Because of cardiac disease, heart muscles stiffen in diastolic heart failure. The left ventricle is not filled with enough blood. A decrease in blood volume in the heart lowers blood flow to the body. This is referred to as systolic heart failure when the heart muscles cannot contract adequately. When the heart beats, oxygenated blood flows from the heart into the body. Systolic dysfunction occurs when the heart does not contract adequately. Inadequate blood supply from the heart to the body is caused by ineffective cardiac contraction [15].

II. RELATED WORK

The researchers' heart disease detection systems were built using machine learning methodologies. Hedeshi and Abadeh used the PSO (particle swarm optimization) technique to derive principles that might be used to identify cardiac disease. In each PSO run, many rules were retrieved. The authors worked on a dataset generated by integrating Cleveland, Long Beach, Hungarian, and Swiss datasets. The dataset includes 920 records, and 13 characteristics were used to diagnose heart disease. A precision of 85.76% was attained. Bashir et al. used the ensemble process to create a system for predicting cardiac disease. The authors ran trials on five different datasets. The absence of illness was indicated by a class label of 0, while the presence of disease was indicated by a class label of 1. The findings were validated using tenfold cross-validation. The authors did not undertake feature selection since the dataset they utilized included only relevant attributes. Missing values and outliers were removed by data preparation. Outliers were identified using the inter-quantile range method[16]. The classification used five classifiers: decision tree induction, information gain, naive Bayes, memory-based learner, support vector machine, and decision tree induction. To improve accuracy, classifier outputs were combined using a majority vote. The Cleveland dataset had an accuracy of 85.81%, the SPECTF dataset had an accuracy of 80.15%, the SPECT dataset had an accuracy of 82.40%, the Eric dataset had an accuracy of 86.12%, and the Statlog dataset had an accuracy of 88.52%. Tomar et al. used LSTSVM to create a system for detecting cardiac problems (least square twin support vector machine). The features were chosen based on their F-score. Experiments were run on the Statlog dataset, and an accuracy of 85.59% was obtained. Olaniyi and Oyedotun created a model for heart disease detection based on MLP and SVM (support vector machine).

For training the MLP, the backpropagation technique was utilized. MLP employed a learning rate of 0.32. Experiments were carried out using the Statlog dataset, which had 270 occurrences. SVM had an accuracy of 87.5%, whereas multilayer perceptron had an accuracy of 85%. Marateb and Goudarzi used a fuzzy rule-based system built by NFC to identify the cardiac disease (Neuro-Fuzzy Classifier). The scaled conjugate gradient approach was utilized to significantly minimize the root mean square error and accelerate NFC learning. The authors in their study utilized the Cleveland dataset. Discretization was used as part of data preparation to convert continuous values of characteristics into discrete values. Fuzzification and defuzzification were used in the categorization process. The SFS (sequential feature selection) and MLR (multiple logistic regression) algorithms were utilized to discover relevant features. The hold-out approach was used to validate the system. NFC performance was evaluated without the use of feature selection, as well as with MLR and SFS. The findings showed that NFC, in conjunction with MLR, had the greatest accuracy of 84%.

III. PROPOSED METHODOLOGY

Machine learning is a current trend in computer science that is being utilized to create a wide range of decision-assistance systems. High-dimensional data is a prevalent issue when decision support systems are utilized in real-world applications. This issue is frequently referred to as the curse of dimensionality [16]. Unnecessary data characteristics have an impact on categorization system performance.

ML Techniques for Heart Disease Prediction

The initial stage in building any decision support system is to pick critical characteristics. Important characteristics are chosen using feature selection techniques. Pick relevant characteristics; many feature selection techniques are available. Each feature selection technique picks a distinct set of features when applied to a collection of features [17-35]. The characteristics chosen have an impact on classification performance. As a result, it is critical to employ an effective feature selection approach to extract significant characteristics from the dataset. This article introduces numerous feature selection approaches. The GARFE (genetic algorithm recursive feature elimination), CFGA (correlation feature selection genetic algorithm), and HPCBE (hybrid Pearson correlation with backward elimination) hybrid feature selection methods are also presented. The suggested feature selection algorithms' performance is compared to various current feature selection methods.

Feature Selection Methods

Methods for selecting relevant characteristics aid in the reduction of the Number of features. The system's accuracy improves as the Number of characteristics is reduced. It also minimizes the system's complexity. Removing superfluous features also contributes to shorter calculation times [18]. This section discusses the many kinds of feature selection strategies and their benefits and drawbacks. There is also an explanation of several regularly used feature selection algorithms. Figure 2 depicts the classification of feature selection approaches.

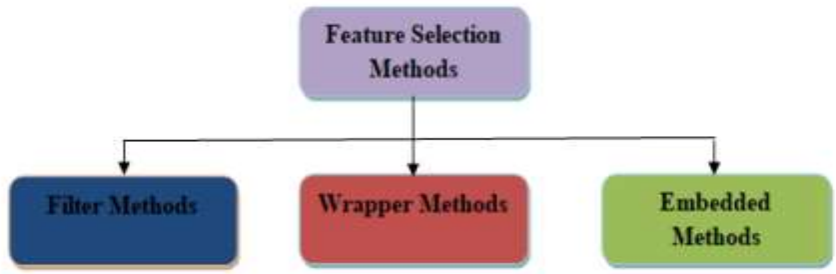

Figure 2: Types of feature selection methods

Before applying the features to the classification algorithm, filter techniques filter them. They score the features based on the data's general qualities. This rating is used to pick the features [19]. The way for selecting features in the filter methods. Wrapper techniques pick features by training the model on a subset of features numerous times and choosing the best subset each time. The model's conclusions are used to choose which feature subset to select. The feature selection and training procedures are combined in embedded approaches. The features are chosen when the model is being trained. From the whole collection of features, several feature subsets are produced. Training and assessing a model on these characteristics determine the effectiveness of these feature subsets.

Proposed Hybrid GARFE Algorithm

Figure 3 shows a novel feature selection method GARFE (genetic algorithm recursive feature elimination), presented in this part to choose relevant features from the dataset. GARFE is a hybrid feature selection method that combines the genetic algorithm with recursive feature elimination, two effective feature selection techniques. The SVM (support vector machine) classifier was used to assess the effect of GARFE on classification performance. Accuracy, sensitivity, specificity, precision, and F-measure were used to assess performance. The method was written in the Python 3.0 programming language. Jupyter notebook 70's integrated development environment was used to do Python experiments. The GARFE method is pitted against eight other feature selection algorithms. The ten-fold cross-validation approach is used to verify the results.

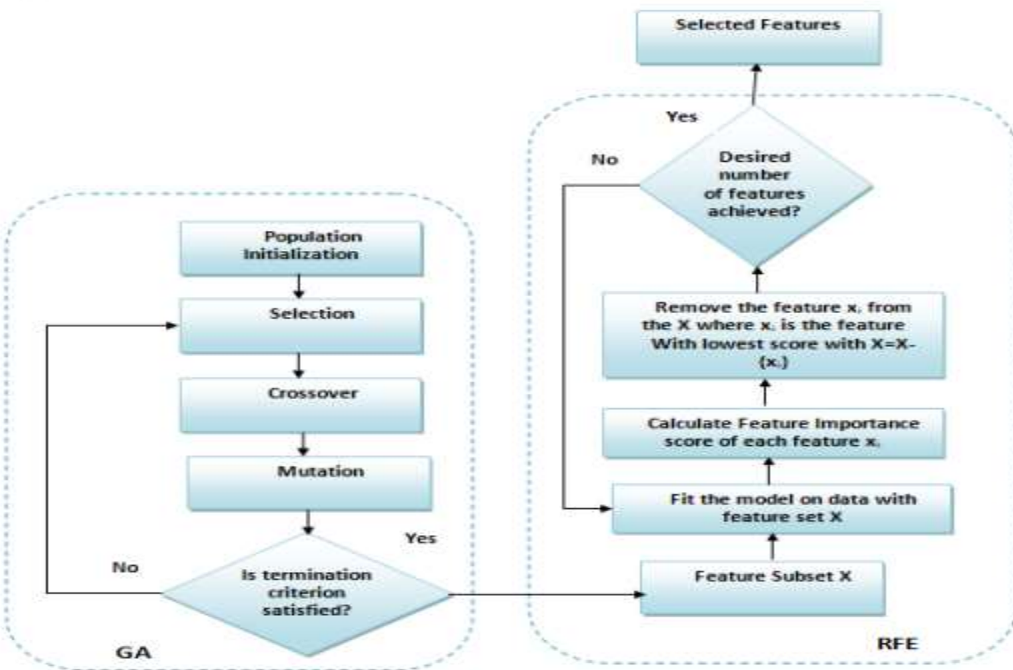

Figure 3. Flowchart of GARFE algorithm

First, a genetic algorithm is used to maximize the collection of characteristics. The optimized set of features is then sent through the recursive feature elimination method, reducing the Number of features.

Pseudocode of GARFE algorithm

```
Algorithm GARFE( )
{
 Number of chromosomes = Population size
 N (Number of genes in chromosome) = Number of features
 i=1
 While ( i <= Number of generations)
 {
 Perform selection to select chromosomes
 Perform crossover to produce new chromosomes
 Perform mutation to bring random changes in chromosomes
 Use fitness function to select population for next generation
 i=i + 1
 }
Best chromosome is produced with N genes.
```

```
Feature Set X={ }
j=1
while (j <= N) //Repeat the steps for each gene in best chromosome
{
if (value of gene j = 1) //if value of gene is 1, feature is selected
{
Add the feature number j to set X
}
}
while(optimum number of features achieved)
{
Use a classifier to fit on data using feature set X.
Calculate the importance score of each feature.
Remove the feature having the lowest score.
}
}
```

The SVM (support vector machine) classifier was used to assess the effect of GARFE on classification performance. Accuracy, sensitivity, specificity, precision, and F-measure were used to assess performance. The method was written in the Python 3.0 programming language. An integrated development environment of Jupyter notebook was employed to conduct Python experiments. The GARFE method is pitted against eight other feature selection algorithms. The ten-fold cross-validation approach is used to verify the results.

IV.RESULTS AND DISCUSSION

The GARFE's performance is compared to eight other feature selection methods. The influence of feature selection strategies on SVM classification performance has been compared. On the Cleveland and Statlog datasets, the performance of the SVM classifier with various feature selection strategies was evaluated. In the Cleveland dataset, the GARFE improved accuracy by 4.07%, sensitivity by 3.14%, specificity by 7.52%, precision by 4.18%, and F-Measure by 3.70%. Figures 3.21–3.25 provide a comparison of the performance of GARFE with different feature selection techniques on the Statlog dataset. Compared to other feature selection algorithms, these results show that GARFE has the greatest accuracy, sensitivity, specificity, precision, and F-Measure. The Statlog dataset improved by 7.16% accuracy, 9.18% sensitivity, 5.95% specificity, 12.18% precision, and 9.91% F-Measure.

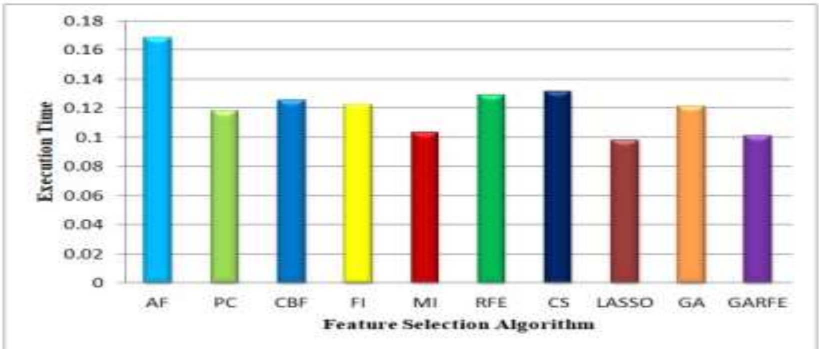

Figure 4: Comparison of execution Time of GARFE with other feature selection methods on Statlog dataset

Figure 4 demonstrates a comparison of execution time in training an SVM classifier with features picked using various feature selection techniques on two datasets. GAREF produced the shortest execution time. It lowered execution time by 48.27% in the Cleveland dataset and 39.88% in the Statlog dataset. In both datasets, the GARFE feature reduction ratio is 38.46%. It produced the quickest reaction time and best accuracy. The GARFE algorithm picked significant characteristics, enhancing classifier performance and minimizing execution time. Figure 4 depicts the distribution of continuous characteristics in the dataset. A tenfold cross-validation procedure is used to verify the results. Various models were built utilizing the CAD dataset and various classifiers such as XGBoost, KNN, SVM-Linear, and SVM-RBF. Classifier performance has been evaluated using accuracy, sensitivity, specificity, precision, and F-Measure measures. In addition to these performance factors, models were tested using ROC curves (receiver operating characteristic curve).

V. CONCLUSION

In today's environment, machine learning has become an essential component of our daily lives. Machine learning suffers from the challenge of high-dimensional data when applied to real-world applications like e-healthcare. Data may include duplicate and extra elements. These superfluous characteristics impact the performance of classification systems used in prediction. Three hybrid feature selection methods are presented in this article. The proposed hybrid algorithms have two advantages: they improve classification accuracy while lowering feature dimensionality. The SVM classifier was used to assess GARFE's performance. The LR classifier was used to assess the performance of CFGA. HPCBE is assessed using four different classifiers: DT, KNN, XGBoost, and AdaBoost. The accuracy, specificity, sensitivity, precision, F-measure, and execution time properties of feature selection algorithms have been evaluated.

REFERENCES

1. David, D. S., Arun, S., Sivaprakash, S., Raja, P. V., Sharma, D. K. et al. (2022). Enhanced Detection of Glaucoma on Ensemble Convolutional Neural Network for Clinical Informatics. CMC-Computers, Materials & Continua, 70(2), 2563–2579.
2. David, D. S., Anam, M., Kaliappan, C., Arun, S., Sharma, D. K. et al. (2022). Cloud Security Service for Identifying Unauthorized User Behaviour. CMC-Computers, Materials & Continua, 70(2), 2581–2600.
3. Jayachandran, A., and D. Stalin David. "Textures and Intensity Histogram Based Retinal Image Classification System Using Hybrid Colour Structure Descriptor." Biomedical and Pharmacology Journal, vol. 11, no. 1, 2018, p. 577+. Accessed 12 Feb. 2021.
4. D. Stalin David, 2019, "Parasagittal Meningiomia Brain Tumor Classification System based on MRI Images and Multi Phase level set Formulation", Biomedical and Pharmacology Journal, Vol.12, issue 2, pp.939-946.
5. Thendral, R., & David, D. S. (2022). An Enhanced Computer Vision Algorithm for Apple Fruit Yield Estimation in an Orchard. In Artificial Intelligence and Technologies (pp. 263-273). Springer, Singapore.
6. D. S. David and A. Jeyachandran, "A comprehensive survey of security mechanisms in healthcare applications," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp. 1-6, doi: 10.1109/CESYS.2016.7889823.
7. Stalin David, D., Jayachandran, A. A new expert system based on hybrid colour and structure descriptor and machine learning algorithms for early glaucoma diagnosis. Multimed Tools Appl 79, 5213–5224 (2020). https://doi.org/10.1007/s11042-018-6265-1.

8. D Stalin David, A Jayachandran, 2018,Robust Classification of Brain Tumor in MRI Images using Salient Structure Descriptor and RBF Kernel-SVM, TAGA Journal of Graphic Technology, Volume 14, Issue 64, pp.718-737.
9. D Stalin David, 2016, Robust Middleware based Framework for the Classification of Cardiac Arrhythmia Diseases by Analyzing Big Data, International Journal on Recent Researches In Science, Engineering & Technology, 2018, Volume 4, Issue 9, pp.118-127.
10. M. Rajdhev, D. Stalin David, "Internet of Things for Health Care", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 800-805, March-April 2017.
11. P. Prasanth, D. Stalin David, "Defensing Online Key detection using Tick Points", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 758-765, March-April 2017.
12. A.Sudalaimani, D. Stalin David, "Efficient Multicast Delivery for Data Redundancy Minimization over Wireless Data Centres", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 751-757, March-April 2017.
13. R. Abish, D. Stalin David, "Detecting Packet Drop Attacks in Wireless Sensor Networks using Bloom Filter", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 730-735, March-April 2017.
14. A.Vignesh, D. Stalin David, "Novel based Intelligent Parking System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 2 Issue 2, pp. 724-729, March-April 2017.
15. D Stalin David, 2020, 'Diagnosis of Alzheimer's Disease Using Principal Component Analysis and Support Vector Machine, International Journal of Pharmaceutical Research, Volume 12, Issue 2, PP.713-724.
16. Jaswanth K S, Dr. D. Stalin David, "A Novel Based 3d Facial Expression Detection Using Recurrent Neural Network", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 48-53, March-April 2020.
17. D Stalin David, 2020, 'An Intellectual Individual Performance Abnormality Discovery System in Civic Surroundings' International Journal of Innovative Technology and Exploring Engineering, Volume 9, Issue 5, PP.2196-2206.
18. D Stalin David, 2020, 'Machine learning for the prelude diagnosis of dementia', International Journal of Pharmaceutical Research, Volume 13, Issue 3, PP.2329-2335.
19. David, D.S. and Y. Justin, 2020.A Comprehensive Review on Partition of the Blood Vessel and Optic Disc in RetinalImages.Artech J. Eff. Res. Eng. Technol., 1: 110-117.
20. D. Stalin David and A.A. Jose, 2020. Retinal image classification system for diagnosis of diabetic retinopathy using SDCMethods.Artech J. Eff. Res. Eng. Technol., 1: 87-93.
21. D. Stalin David and T. Joseph George, 2020. Identity-based Sybil attack detection and localization.ArtechJ. Eff. Res. Eng. Technol., 1: 94-98.
22. David, D.S. and L. Arun, 2020.Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme.Artech J. Eff. Res. Eng. Technol., 1: 57-63.
23. David, D.S., 2020. Retinal image classification system for diagnosis of diabetic retinopathy using morphological edgedetection and feature extraction techniques.Artech J. Eff. Res.Eng. Technol., 1: 28-33.
24. David, D.S., 2020. A novel specialist system based on hybrid colour and structure descriptor and machine learningalgorithms for early diabetic retinopathy diagnosis.Artech J. Eff. Res. Eng. Technol., 1: 50-56.
25. David, D.S. and M. Samraj, 2020.A comprehensive survey of emotion recognition system in facial expression.Artech J. Eff. Res. Eng. Technol., 1: 76-81.
26. David, D.S. and L. Arun, 2020. Multi-view 3D face renovation with deep recurrent neural networks. ArtechJ. Eff. Res. Eng. Technol., 1: 64-68.
27. David, D.S. and S. Namboodiri, 2020.Improvement of framework for the grouping of CA diseases by investigating bigdata.Artech J. Eng. Appl. Technol., 1: 7-14.
28. Stalin David D , Saravanan M, 2020, 'Multi-perspective DOS Attack Detection Framework for Reliable Data Transmission in Wireless Sensor Networks based on Trust', International Journal of Future Generation Communication and Networking , Volume 13, Issue 4, PP.1522–1539.
29. J. K. S and D. S. David, "A Novel Based 3D Facial Expression Detection Using Recurrent Neural Network," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-6, doi: 10.1109/ICSCAN49426.2020.9262287.
30. Stalin David D, Saravanan M, "Enhanced Glaucoma Detection Using Ensemble based CNN and Spatially Based Ellipse Fitting Curve Model", Solid State Technology, Volume 63, Issue 6, PP.3581-3598.
31. Stalin David D, Saravanan M, Jayachandran A, "Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification", Solid State Technology, Volume 63, Issue 6, PP.3599-3623.
32. D. Jayakumar; Dr.U. Palani; D. Raghuraman; Dr.D. StalinDavid; D. Saravanan; R. Parthiban; S. Usharani. "CERTAIN INVESTIGATION ON MONITORING THE LOAD OF SHORT DISTANCE ORIENTEERING SPORTS ON CAMPUS BASED ON EMBEDDED SYSTEM ACCELERATION SENSOR". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2477-2494.
33. R. Parthiban; S. Usharani; D. Saravanan; D. Jayakumar; Dr.U. Palani; Dr.D. StalinDavid; D. Raghuraman. "PROGNOSIS OF CHRONIC KIDNEY DISEASE (CKD) USING HYBRID FILTER WRAPPER EMBEDDED FEATURE SELECTION METHOD". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2511-2530.

34. Dr.U. Palani; D. Raghuraman; Dr.D. StalinDavid; R. Parthiban; S. Usharani; D. Jayakumar; D. Saravanan. "AN ENERGY-EFFICIENT TRUST BASED SECURE DATA SCHEME IN WIRELESS SENSOR NETWORKS". European Journal of Molecular & Clinical Medicine, 7, 9, 2021, 2495-2510.
35. David, D. D. S., Parthiban, R., Jayakumar, D., Usharani, S., RaghuRaman, D., Saravanan, D., & Palani, D. U. (2021). Medical wireless sensor network coverage and clinical application of MRI liver disease diagnosis. European Journal of Molecular & Clinical Medicine, 7(9), 2559-2571.