ISSN: 2666-2795 Vol. 6 No.1 June, 2021, Netherland

International Journal of Applied Engineering Research

Prediction of Covid-19 Infected Chest X-Rays Using Machine Learning Techniques

[1] Ch Hima Bindu, [2]Maruturi Haribabu

[1]Professor& HOD, ECE Department QIS College of Engineering and Technology, Ongole, India

E-mail Id: ecehod@qiscet.edu.in

[2]Assistant Professor, ECE Department QIS College of Engineering and Technology, Ongole, India

E-mail Id: haribabu.m@qiscet.edu.in

Abstract- As the number of cases with COVID-19 continues to climb, better medical screening and clinical care of the condition are urgently needed. Typical signs of a chest cold include a sore throat, cough, and a high temperature. Patients with pneumonia also have these symptoms. Because of this, detecting COVID-19 is much more difficult. Recognizing COVID-19 in a set of Chest-X-Ray (CXR) pictures that also contains pneumonia patients is laborious and prone to human mistake. Radiography of the chest for COVID-19 cases and others with comparable symptoms may be used as a first-line triage method. While this is true, radiologists still have a hard time distinguishing between COVID-19CXR pneumonia and other types of pneumonia due to the similarities in their appearance. This work is an effort to construct a machine learning model that is beneficial in categorizing CXR pictures into three classes indicating normal, COVID-9, and pneumonia based on the premise that such classifiers can consistently identify COVID-19 CXR images from other kinds of pneumonia. Feature The methods of extraction, dimensionality reduction, and machine learning are all used.

Index Terms- Chest X-Rays, Covid-19, MachineLearning, Prediction, Pneumonia.

I. INTRODUCTION

In December of 2019, a corona virus disease (COVID - 19) pandemic began in Wuhan, China and quickly spread over the globe. There is currently no cure or vaccination available for COVID -19. Severe acute respiratory syndrome corona virus (also known as COVID-19) is the causative agent of the current COVID-19 pandemic. Worldwide, - 19) has infected millions and killed a number of people. The world's biggest issue is COVID-19. The epidemic has been labeled a "global health emergency" by the World Health Organization (WHO). The lungs are particularly vulnerable to Corona virus illness. Temperature elevation, dehydration, coughing, and other respiratory distress are among Corona virus symptoms. As a newly identified species in 2019, COVID-19 has not been tentatively connected to humans. Early evidence suggests that the symptoms caused by COVID-19 are milder than those caused by hepatitis C virus (HCV), which may range from moderate to life-threatening. Faster triage of non-COVID-19 patients and more efficient use of hospital resources for COVID-19 cases are both possible thanks to an accurate automated system for classifying CXR pictures according to the presence or absence of the COVID-19 virus. The reliable interpretation of medical pictures has been greatly aided by technologies founded in Machine Learning (ML). Methods that are based on machine learning are able to scale, be automated, and be easily used in practical situations. A typical task in ML-based image analysis is the categorization of photos having similar properties. To directly assign picture samples into target classes, this strategy focuses on segmenting the region of interest, identifying useful image features extracted from the segmented area in the spatial or frequency domain, and developing an optimum machine learning-based classification technique. Triage of individuals with comparable symptoms but not COVID-19 might begin with a chest X-ray (CXR). Classifiers trained with machine learning can consistently identify COVID-19 patients in CXR pictures by separating them from patients with other types of pneumonia. Feature extraction, a method of digital image processing, and dimensionality reduction, a method of machine learning, are used to develop a reliable machine learning classifier in this study, one that can accurately and sensitively identify COVID-19 cases from those without the virus. Literature review, suggested study, findings, and a conclusion make up the remainder of the article.

II. LITERATURESURVEY

New machine learning-based approaches for analysing COVID-19 medical pictures have been proposed. CT scans were used by Wang et al.[2] to construct a deep learning network, DenseNet121, that had previously been taught to categories using data from the COVID-19 imaging tests. In determining whether the data was positive or negative, 81.24 percent of the time, the prediction was correct. Prevention relies on identifying high-risk patients and putting available medical resources to the best use, and early diagnosis of COVID-19 is essential for this. Zhang et al. [3] recommended training on ResNet-18 for improved lung lesion segmentation in CT images. The accuracy of the classifier model is 92.49 percent for the combined pneumonia and normal COVID-19 categories. The study's goal was to determine whether or not combining spatial and frequency-domain imaging characteristics may better characterise tumour heterogeneity and predict tumour response to postoperative chemotherapy in patients with advanced-stage ovarian cancer. To begin, a computer-aided method was used to calculate 133 attributes. The characteristics were categorised into shape/density, fast Fourier transform, wavelet, discrete cosine transform (DCT), and grey level difference. The method relied on the particle swarm optimization technique to create a cluster of characteristics with the right amount of variation to accomplish better discrimination than was possible using any of the features individually.

Clinical data from 120 ovarian cancer patients was collected in a retrospective manner and used to evaluate the methodology. DCT features showed the highest prediction accuracy across all five groups when comparing individual feature findings. The best feature cluster produced a quantitative image marker with an area under the ROC curve (AUC) of 0.86, while the best single feature produced an AUC of just 0.74. Properties in the frequency domain were also shown to be important, alongside those in the spatial domain. In conclusion, this study's findings reveal that a combination of the proposed new quantitative image marker and features determined from the spatial and frequency domains might effectively predict the response of a tumour to chemotherapy after surgical removal.

CT imaging features of his novel virus are shown to be distinct from those of previously described viral pneumonias, as determined by Xuet al. They use a number of different CNN models for image classification in computed tomography to help in the early identification of COVID-19. A total of 618 CT images are obtained, from 175 healthy people, 219 COVID-19 infected patients, and 224 CT samples of influenza A virus pneumonia. They use a three-dimensional convolutional neural network (CNN) model based on the well-tested ResNet-18 network architecture to partition the possible image cubes. The author makes use of a system for classifying images in three dimensions in order to arrange all of the image blocks in the text. This categorization scheme was developed to help make sense of the connections seen in medical imaging of the lungs. The overall confidence score of CT cases and the specific infection type (COVID-19, influenza A viral pneumonia, or no infection diagnosed) are both calculated using noisy or Bayesian functions.

Shan et al. [9] propose a deep learning system they name "VB Net" to automatically segment and measure the corona virus-infected regions in computed tomography images. The VB Net model integrates features from both the V-Net and bottleneck models. While both the V-Net and the bottleneck model utilise convolution and down sampling to pull out broad picture characteristics, the bottleneck model further use convolution and up sampling to blend in more granular information. First, 249 people with COVID-19 were used to train the model, and then another 300 COVID-19 patients were used to verify it. To expedite the laborious task of displaying COVID-19 CT scans for training, the author offers a "in loop (HITL)" technique that iteratively generates training examples. After three iterations, the drawing time is reduced to 4 minutes using the proposed manual loop approach for updating the model. Using the Dice correspondence coefficient, we find that the deep learning model achieves 91.6% accuracy in subdividing the whole 300-verification-set.

III. PROPOSEDMETHOD

This project work carrries the following steps to build a model that classifies the CXR image of a person who is either healthy orinfected by COVID-19 or Pneumonia which is shown in Fig.1. It involves Collecting datasets, Data preprocessing, Feature extraction,Featureevaluation,Featurereduction, Modeltrainingandtesting.

ALGORITHM

Step 1:CollectingDatasets

Acollectionof420CXRimages[11,12]storedinthreefoldersnamely normal(indicatingperfectlyhealthyperson),covid(indicatingcovidinfectedperson)andpneumonia(indicatingpneumoniainfectedperson)respectively.

STEP2:DATAPREPROCESSING

Datapreprocessinginvolvestwosteps.Min-MaxnormalizationandAdaptiveHistogramequalization.Thesetwostepsareappliedforeachimage.

STEP3:FEATUREEXTRACTION

Five different types of features are extracted for each image, three types of spatial domain features and two types of frequency domainfeatures. Spatial domain features are Texture, GLDM, GLCM and frequency domain features are FFT and Wavelet. The computedfeatures arestoredina.matfileasa420x252arraymeaning252featuresfor420imageseach.

STEP4:FEATUREEVALUATION

Thecomputedfeaturesarethenevaluatedbycalculatingcorrelation matrixandhistogram.

STEP5:FEATUREREDUCTION

Thefeaturesarereducedfrom252to64featureseachusingKernel-PCAmethod.

STEP 6:MODELTRAININGANDTESTING

Adeeplearningmodelwithtwohiddenlayersof128and16dimensionsrespectivelyisbuilt.Thedataisloadedintothemodeland60%isusedfortr ainingand20%fortestingand20%forvalidation.

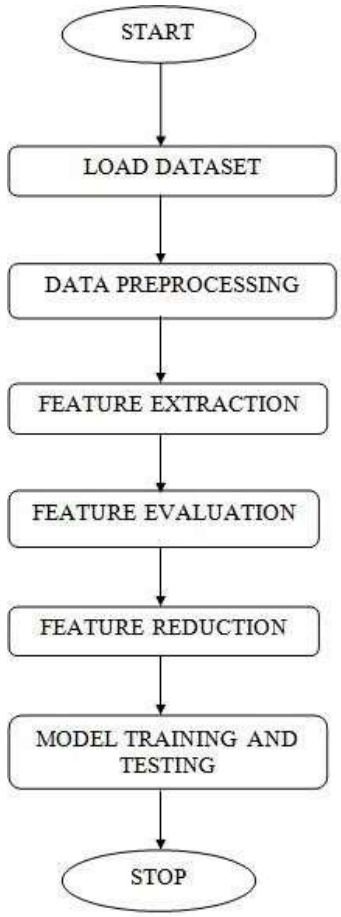

Fig.1AlgorithmFlowchart

V. RESULTSAND DISCUSSION

Theresultsobtainedare shownfromFig.2toFig.5.

CXR of a healthy person (Before)

CXR of a healthy person (After)

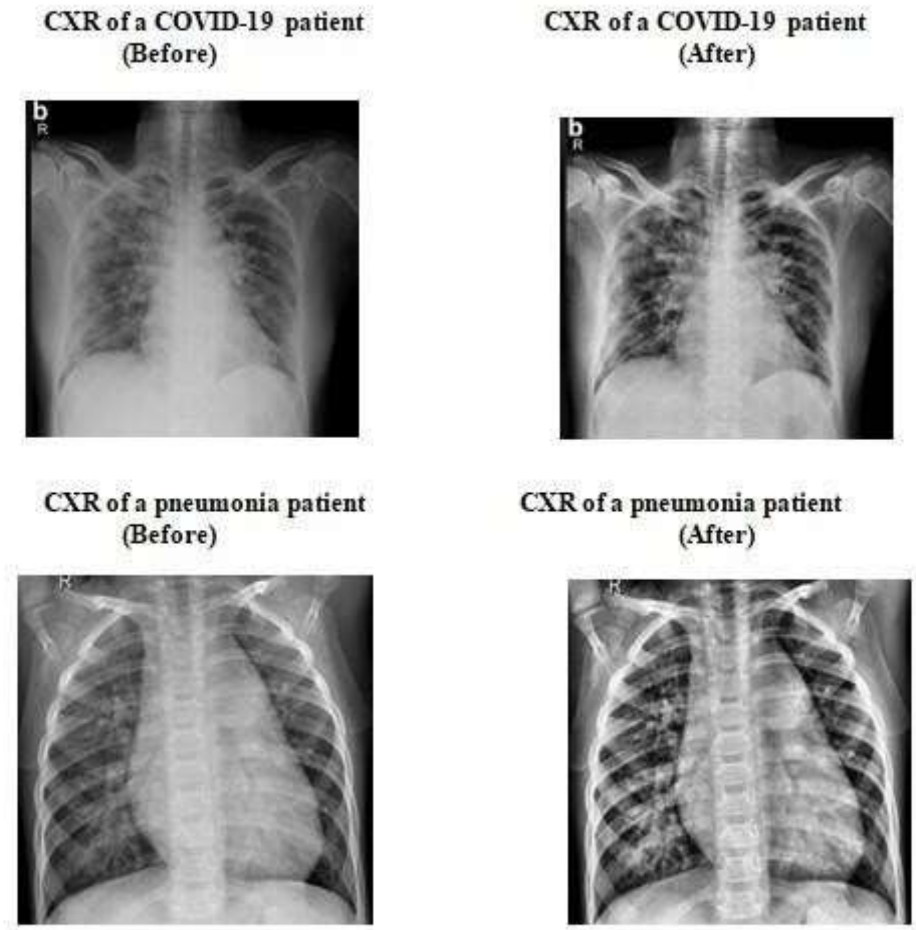

Fig.2OutputofDataPreprocessing

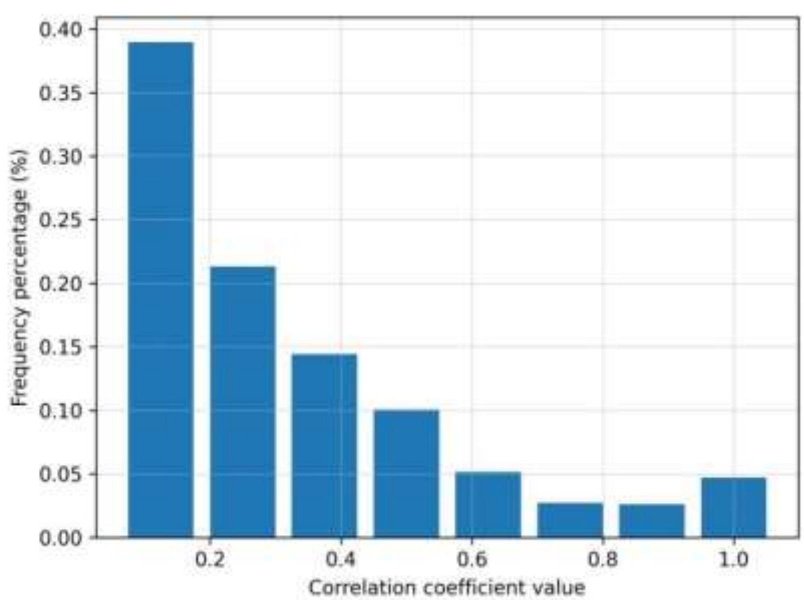

Fig.3Histogramrepresentationofcorrelationcoefficients

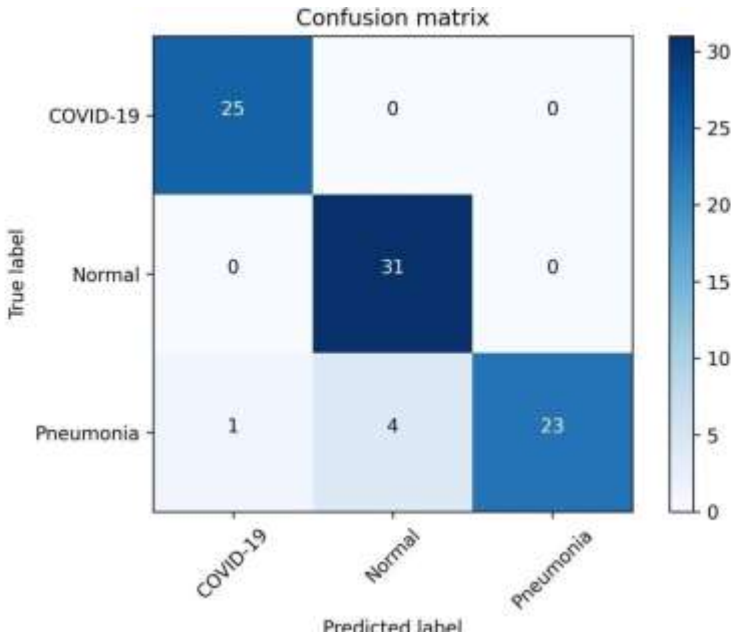

Fig.4ConfusionMatrix

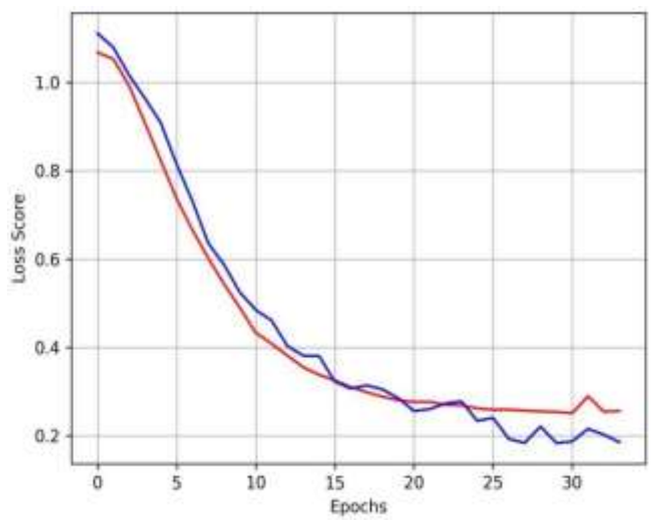

Fig.5Graph comparingtrainingloss(red line)andvalidation loss(blueline)

VI. CONCLUSION

A Model that classifies the Chest-X-Ray images into three classes namely covid, normal and pneumonia isbuilt. Feature extractionand dimensionality reduction techniques are used. A relatively small dataset of 420 Chest-X-Ray images is considered. The overallaccuracy of the model obtained is 94.05%. Machine learning models surely cannot replace the traditional and more accurate diagnosissystems, but the technological advancements happening everyday are making the ML models more and more accurate and reliable.This work attempts to build a reasonably accurate ML model that can be handy while performing the diagnosis of COVID-19 patients.Although the main limitation being consideration of a small dataset, this can be used as a reference to explore any other methods in constructing accurate and reliable Machine Learning models in the future.

VII. ACKNOWLEDGEMENT

The authors would like to express gratitude to the QIS Management.

VIII. REFERENCES

[1] Oh,Y.,Park,S.&Ye,J.C.DeeplearningCOVID-19featuresonCXRusinglimitedtrainingdatasets.IEEETrans.Med.Imaging39,2688–2700.https://doi.org/10.1109/TMI.2020.2993291(2020).

[2] Wang,S.etal.AFullyautomaticdeeplearningsystemforCOVID-19diagnosticandprognosticanalysis.Eur.Respir.J.2000775,https://doi.org/10.1183/13993003.00775-2020 (2020).

[3] Zhang,W.,Xue,X.,Lu,H.&Guo,Y.-F.Discriminantneighborhoodembeddingforclassifcation.PatternRecogn.39,2240–2243.https://doi.org/10.1016/j.patcog.2006.05.011(2006)

[4] Roy,S.etal.DeeplearningforclassificationandlocalizationofCOVID-19markersinpoint-of-carelungultrasound.IEEETrans.Med.Imaging39,2676–2687.https://doi.org/10.1109/TMI.2020.2994459 (2020).

[5] WorldHealthOrganization.ChestRadiographyinTuberculosisDetection.(WorldHealthOrganization,2016).

[6] Zargari,A.etal.Predictionofchemotherapyresponseinovariancancerpatientsusinganewclusteredquantitativeimagemarker.Phys.Med.Biol.63,155020.https://doi.org/10.1088/1361-6560/aad3ab(2018).

[7] Ahmed,Z.,Mohamed,K.,Zeeshan,S.& Dong,X.Artificialintelligencewithmulti-functionalmachine learningplatform developmentforbetterhealthcareandprecision medicine.Database(Oxford)2020,baaa010,https://doi.org/10.1093/database/baaa010 (2020).

[8] Shah,P.etal.Artificialintelligenceandmachinelearninginclinicaldevelopment:Atranslationalperspective.NPJDigit.Med.2,69–69https://doi.org/10.1038/s41746-019-0148-3(2019).

[9] Sun,L.,Shao,W.,Wang,M.,Zhang,D.&Liu,M.High-orderfeaturelearningformulti-atlasbasedlabelfusion:ApplicationtobrainsegmentationwithMRI. IEEETrans.ImageProcess. 29,2702–2713.https://doi.org/10.1109/TIP.2019.2952079 (2020).

[10] Kermany, D.Z., Kang, Gold baum, M .Labeled optical coherence tomography (OCT)andchestX-rayimagesforclassifcation.MendeleyData,v2,https://doi.org/10.17632/RSCBJBR9SJ.2(2018)

[11] Cohen,J.P.,Morrison,P.&Dao,L.COVID-19ImageDataCollection.:2003.11597(2020).

[12] He,K.,Zhang,X.,Ren,S.&Sun,J.DeepResidualLearningforImageRecognition.arXiv:1512.03385(2015).

[13] Kesim,E.,Dokur,Z.&Olmez,T.X-Raychestimageclassifcationbyasmall-sizedconvolutionalneuralnetwork.InScientifcMeetingonElectrical-Electronics&BiomedicalEngineeringandComputerScience(EBBT).1–5,https://doi.org/10.1109/EBBT.2019.8742050 (2019).

[14] Chen,X.&Jeong,J.C.Enhancedrecursivefeatureelimination.InSixthInternationalConferenceonMachineLearningandApplications(ICMLA2007).429–435 (2007).

[15] AbolfazlZargariKhuzani,MortezaHeidari,AliShariati,"COVID-Classifier:Anautomatedmachinelearningmodeltoassistinthediagnosisof COVID-19infectionin chestx-ray images,"medRxiv,doi:https://doi.org/10.1101/2020.05.09.20096560,2020.

[16] Rajkovic, N., Ciric, J., Milosevic, N. &Saponjic, J. Novel application of the gray-level co-occurrence matrix analysis in the parvalbumin stained hippocampalgyrusdentatusindistinctratmodelsofParkinson'sdisease.Comput.Biol.Med.115,103482.https://doi.org/10.1016/j.compbiomed.2019.103482(2019).

[17] Moallem,P.,Zargari,A.&Kiyoumarsi,A.ImprovementincomputationofΔV10Flickerseverityindexusingintelligentmethods.J.PowerElectron.11,228–236.https://doi.org/10.6113/JPE.2011.11.2.228 (2011).

[18] Danala,G.etal.ApplyingquantitativeCTimagefeatureanalysistopredictresponseofovariancancerpatientstochemotherapy.Acad.Radiol.24,1233–1239.https://doi.org/10.1016/j.acra.2017.04.014 (2017).

19. G.Sabarmathi , Dr.R.Chinnaiyan (2018), "Envisagation and Analysis of Mosquito Borne Fevers – A Health Monitoring System by Envisagative Computing using Big Data Analytics" in ICCBI 2018 – Springer on 19.12.2018 to 20.12.2018 (Recommended for Scopus Indexed Publication IEEE Xplore digital library)
20. G.Sabarmathi , Dr.R.Chinnaiyan, Reliable Data Mining Tasks and Techniques for Industrial Applications, IAETSD JOURNAL FOR ADVANCED RESEARCH IN APPLIED SCIENCES, VOLUME 4, ISSUE 7, DEC/2017,PP- 138-142, ISSN NO: 2394-8442
21. Dr. M. Thangamani, Jafar Ali Ibrahim, Information Technology E-Service Management System, International Scientific Global Journal in Engineering Science and Applied Research (ISGJESAR). Vol.1. Issue 4, pp. 13-18, 2017. http://isgjesar.com/Papers/Volume1,Issue4/paper2.pdf
22. Ibrahim, Mr S. Jafar Ali, K. Singaraj, P. Jebaroopan, and S. A. Sheikfareed. "Android Based Robot for Industrial Application." International Journal of Engineering Research & Technology 3, no. 3 (2014).
23. Ibrahim, S. Jafar Ali, and M. Thangamani. "Momentous Innovations in the Prospective Method of Drug Development." In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, pp. 37-41. 2018.
24. Ibrahim, S. Jafar Ali, and M. Thangamani. "Prediction of Novel Drugs and Diseases for Hepatocellular Carcinoma Based on Multi-Source Simulated Annealing Based Random Walk." Journal of medical systems 42, no. 10 (2018): 188. https://doi.org/10.1007/s10916-018-1038-y ISSN 1311-8080, https://acadpubl.eu/hub/2018-119-16/1/94.pdf
25. Jafar Ali Ibrahim. S, Mohamed Affir. A "Effective Scheduling of Jobs Using Reallocation of Resources Along With Best Fit Strategy and Priority", International Journal of Science Engineering and Advanced Technology(IJSEAT) – ISSN No: 2321-6905, Vol.2, Issue.2, Feb-2014, http://www.ijseat.com/index.php/ijseat/article/view/62
26. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352-357 http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf
27. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352-357 http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf
28. S. Jafar Ali Ibrahim and M. Thangamani. 2018. Momentous Innovations in the Prospective Method of Drug Development. In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing (DMIP '18). Association for Computing Machinery, New York, NY, USA, 37–41. https://doi.org/10.1145/3299852.3299854
29. S. Jafar Ali Ibrahim and Thangamani, M "Proliferators and Inhibitors Of Hepatocellular Carcinoma", International Journal of Pure and Applied Mathematics (IJPAM) Special Issue of Mathematical Modelling of Engineering Problems Vol 119 Issue. 15. July 2018
30. Thangamani, M., and S. Jafar Ali Ibrahim. "Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso-Wc) and Gene Ontology for Microarray Gene Expression." In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, pp. 48-55. 2018. https://dl.acm.org/doi/abs/10.1145/3299852.3299866
31. Testing" , IEEE International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 512 – 517
32. Dr.R.Chinnaiyan, Abishek Kumar(2017) ,"Construction of Estimated Level Based Balanced Binary Search Tree", 2017 IEEE International Conference on Electronics,Communication, and Aerospace Technology (ICECA 2017), 344 - 348, 978-1-5090-5686-6.
33. R.Chinnaiyan, S.Somasundaram (2012) , Reliability Estimation Model for Software Components using CEP", International Journal of Mechanical and Industrial Engineering (IJMIE) , ISSN No.2231-6477, Volume-2, Issue-2, 2012, pp.89-93.
34. R.Chinnaiyan, S. Somasundaram (2011) ,"An SMS based Failure Maintenance and Reliability Management of Component Based Software Systems", European Journal of Scientific Research, Vol. 59 Issue 1, 9/1/2011, pp.123 (cited in EBSCO, Impact Factor: 0.045)
35. R.Chinnaiyan, S.Somasundaram(2011), "An Experimental Study on Reliability Estimation of GNU Compiler Components - A Review", International Journal of Computer Applications, Vol.25, No.3, July 2011, pp.13-16. (Impact Factor: 0.814)
36. R.Chinnaiyan, S.Somasundaram(2010) "Evaluating the Reliability of Component Based Software Systems " ,International Journal of Quality and Reliability Management , Vol. 27, No. 1., pp. 78-88 (Impact Factor: 0.406)
37. Dr.R.Chinnaiyan, Abishek Kumar(2017), Estimation of Optimal Path in Wireless Sensor Networks based on Adjancy List, 2017 IEEE International Conference on Telecommunication,Power Analysis and Computing Techniques (ICTPACT2017) ,6,7,8th April 2017,IEEE 978-1-5090-3381-2.