Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

Identification of Breast Cancer from Thermal Imaging using SVM and Random Forest Method

Lakshman K [1], Siddharth B. Dabhade[2], Y. S. Rode[3], Karan Dabhade[4], S. Deshmukh[5], Ranjan Maheshwari[6]

[1,4] *NIELIT Aurangabad, Dr. B. A. M. U. Campus, Aurangabad (MS), India*

[2] *Dr. G. Y. Pathrikar College of Computer Science and IT, MGM University, Aurangabad (MS)*

[3] *Jijamata Mahavidyalaya, Buldhana, MS, India*

[5] *UDCSIT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India*

[6] *RTU, Kota (RJ), India*

[1]lakshman.korra@gmail.com, [2]dabhade.siddharth@gmail.com, [3]ys.rode@gmail.com

2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-1571-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOEI51242.2021.9452809

Abstract—**In the recent years, the carcinoma disease in women has significantly observed and it has also caused deaths as per WHO 2018 report of carcinoma statistics recorded approximately 2 lakhs registered cases and around 90000 reported deaths. The speed of survival has become very difficult at higher stages of growth and quite 45% of women's in India suffer from stage 3 and 4 of carcinoma. The target of this research is to deliver a report on carcinoma on the basis of the performance of Support Vector Machine [SVM] methodology and random forest using 5 folds, 10 folds, 20 folds with a training set size 50, 60, 70, 80 and 90 respectively. These techniques have achieved an accuracy of 94.5% and 98.40% through the cross validation of Support Vector Machine [SVM] and Random Forest [RF] method.**

***Keywords*—Thermography, preprocessing, SVM classifier, Random Forest, Breast, Cancer, Malignancy**

I. INTRODUCTION

The X-ray diagnosis approach is known as Mammography which is of low powered was used for early stage detection and diagnosis of cancer using Support vector machine (SVM) there are two main approaches initially to detect tumors having suspicious regions & weak contrast in background and how to extract features which will categorize tumours[1]. Hence SVM is meant for classification and a good method of statistical learning has made significant achievement in various fields. This method describes the way to detect of tumour in mammogram images with than 85% of accuracy using SVM.
Objective of this paper is to identify differentiate benign breast tumour from malignant by computer aided diagnostic system two stages involved in RF initial step is backward elimination which is used for feature selection and next step is using random forest are hybridized and in first stage data reduction process leads to reduction in which gives fast results. Our proposed method has averagely obtained 98.4% so, here, both SVM and Random forest techniques are compared to show the result obtained and shown the results, which are compared among SVM methods and Random Forest methodology, where the graphs and values are covered in this paper.

II. LITERATURE SURVEY

This analysis created at terribly early son detection of growth of port cancer. S. Thamarai detection & work focuses on 2 things: characteristic region of growth with minimum distinction conjointly methodology to get the options from these regions to reason tumors Jawad conifer, X-ray photograph segmentation supported automatic technique. The analysis of SVM classifier has been thought-about. First of all, tumors are detected from mammograms with the assistance of morphological process of breast pictures. The GLRLM associated distinction of Gaussian (DoG) options that are extracted from the chosen region, where these options are utilized by SVM; the rule has achieved an accuracy of eighty nine.11% mistreatment SVM classifier. The Y. Ireaneus Pakistani monetary unit Rejani et. al. had planned system that focuses on the solution[3]. Hence to find tumors with a weak distinction at background as a suspicious region and another could be a thanks to extract options that reason tumors. The growth detection methodology follows the theme of (I) sweetening of X-ray photograph (II) The growth space segmentation (III) The extraction of options from the metameric growth space. (IV) Use of SVM classifier could also be outlined as conversion of the image quality to a stronger and additional perceivable level [4][5]. SVM i.e. Support Vector Machine was introduced by Vapnik & his colleagues in 1990's that well-tried to effective & promising techniques for process (Peng et al. 2008 Yang, Wu 2006).

III. PROPOSED METHODOLOGY

To obtain good accuracy in results, research work has been done in SVM or random forest algorithm.

A. Random forest

The random forest belongs to family of classification methods. The following main steps describe construction of RF.

1. From obtained raw or natural data we draw a n tree bootstrap samples.
2. Second step is to grow tree and for each data set and node, a select m-try variables randomly used for

splitting the tree in order that every single terminal should have no fewer than node size cases.

3. Third step is for brand spanking new data prediction the combination information of n tree is completed.
4. Fourth step is to use data which isn't in bootstrap sample to calculate OOB error rate. With **such a lot of** missing values RF can handle no of variables.

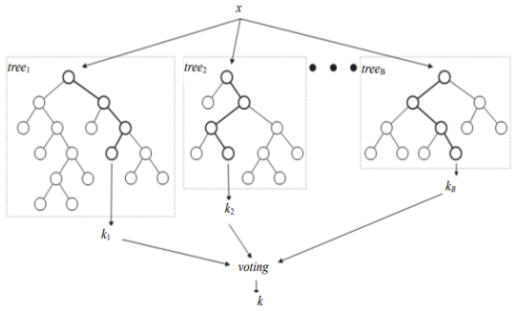

Fig. 1 Random Forest Architecture

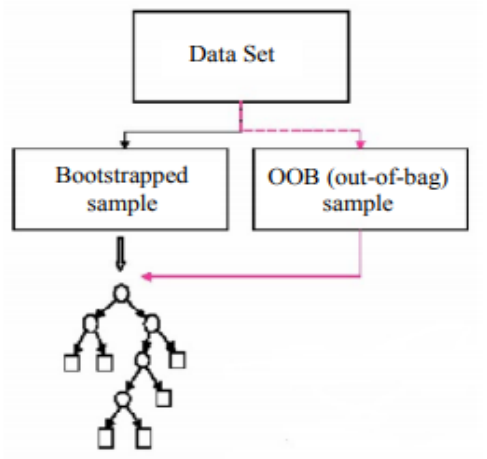

Fig. 2 Random forest sample and out of bag n

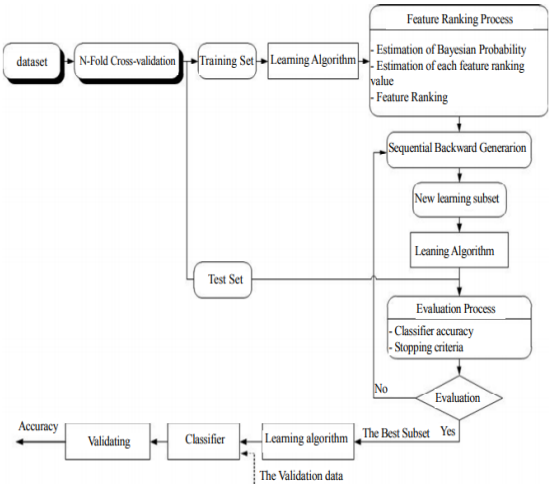

Fig. 3 Existing RF methods

The OOB data is used as test set, test set error estimate by RF and also for importance of variables these two estimates are useful elements for RF out of four the two variable measures which are of node impurity and classification accuracy for OOB data on GINI index [15].

$$G(t) = 1 - \sum_{k=1}^{Q} p^2(k \mid t) \text{ ---------- (1)}$$

Where,'t' is node
P is estimated class probabilities
This can be 2 phase method within which the phase one is of learning algorithm RF were trained & so as to urge the simplest features
To evaluate contribution for features to classifier backward elimination were applied. So on improve the classification accuracy
The Four-step classification algorithm as following:
1: Employment of n-fold crosses validation
2: The Feature ranking value estimate and rank the features.
3 Begin with tiniest feature within the feature ranking list and eliminate features use backward elimination approach
4: Verify the stop criteria;
And jump to Step 1 if not else stop the strategy.

B. Support Vector Machine

The three essential elements of svm like the principle of maximal margin, dual theory, and kernel trick makes SVM successful
Neural Networks having the underside in (SLT) and also the optimization methods are considered as powerful tools to resolve the problems of machine learning.
(n-1) dimension is hyper plane. In higher dimension hyper plane isn't visualized but the notion of an (n-1).

To maximum-margin hyper planes to get nonlinear classifier P-dimensions a kernel trick need to be applied, a hyper plane is described as follows.
The hypothetical values are β0, β1; β2 and Xp are the Data points in sample space.

Feature Selection via SVM:

It's important in machine learning is features of subset which contribute to classification The two fold is benefits of feature selection can leads expensive models which are usually used in many scientific problems and provides classification accuracy.
Tuning of several parameters is key success of Support vector machine which may affect the generalization error which is considered as effective approach.
As estimator is reduces choosing these parameters is to estimate the generalization error and then search for parameters.
While which will generalize well SVMs have been successfully applied economics, finance and management and various fields.
In financial forecasting problems there are applications of SVM which have been reported. Some applications in the field of economics including environmental risk assessment e-banking website quality assessment can also be explored by SVM.

IV. Experimental Work

The Wisconsin Diagnosis carcinoma is dataset obtained from (UCI) an repository of Machine Learning Repository which is usually employed by researchers for purpose of assorted methods in field of machine learning methods for carcinoma Wisconsin carcinoma (WBC) and name of Brest cancer whose data obtained by UCI dataset [11].

Around 1000 instances were available during this dataset where the cases are labelled as either benign or malignant and 458 (65.50%) of the cases are benign and 241 (34.50%) which are found malignant & hence the partition of dataset is done into two parts referred as classes here 2 denotes the benign class and 4 denotes the malignant class [17].

Performed cross-validation method on 5, 10 and 20 fold on 50,60,70,90,training set we've got compared the results of both SVM classifier and RF algorithm to indicate the most effective approach with accuracy stance here an odd number is chosen as datum if the amount of classes is 2.

TABLE I
Comparison between SVM and RF Classifier

Classifier	5 Fold Cross validation				
	50	60	70	80	90
Support Vector Machine	0.878	0.889	0.915	0.939	0.944
Random Forest	0.956	0.952	0.973	0.974	0.979
10 Fold Cross validation					
Support Vector Machine	0.879	0.884	0.912	0.944	0.937
Random Forest	0.951	0.947	0.971	0.979	0.984
20 Fold Cross validation					
Support Vector Machine	0.879	0.89	0.906	0.945	0.94
Random Forest	0.962	0.950	0.974	0.972	0.974

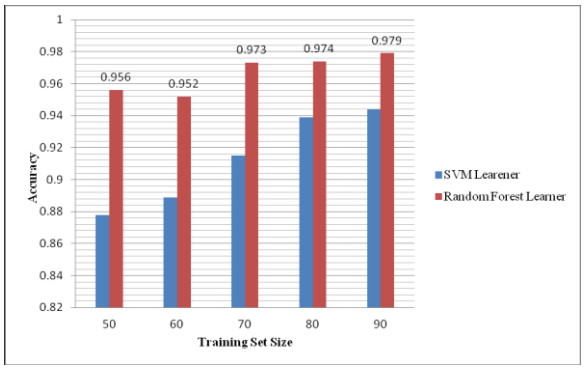

Fig. 4 Comparison SVM and RF with 5 Fold

Then compared the results of Support vector Machine Classifier and Random forest Classifier in proposed model we found that Precision and Recall columns shown below of SVM is lowered value on compared with the RF.

In Table I we compared Support vector Machine Classifier & Random Forest Classifier with 5,10 and 20 fold dataset to show result in SVM results on compared with RF, is Low in each of 5 fold 10 & 20 fold table as seen.

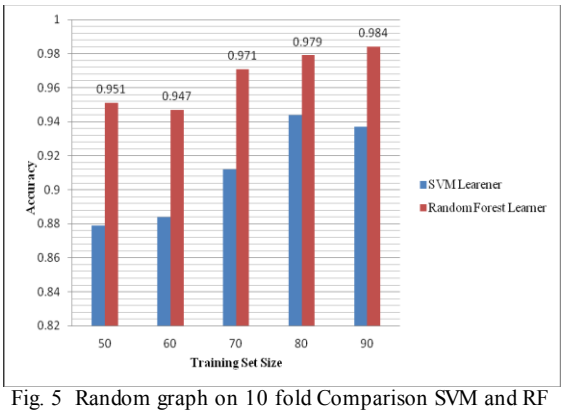

Fig. 5 Random graph on 10 fold Comparison SVM and RF

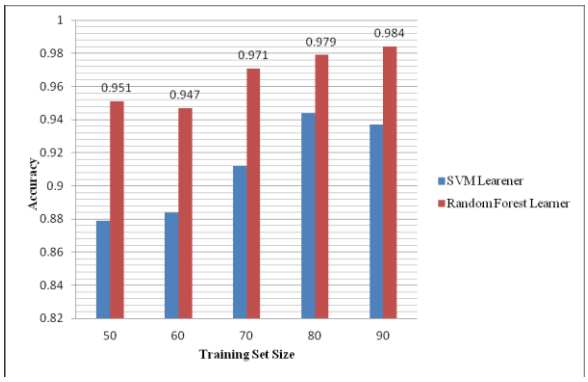

Fig. 6 Random Graph on 20 fold Comparison SVM and RF

Comparison of both methods results are shown in the Table 1. Our proposed paper focus on performance of the SVM learner and Random forest Learner methods among 5 fold, 10 fold and 20 folds each with 50, 60, 70, 80 & 90 training dataset size. As shown in the tables above it is observed and analyzed that the random forest learner method gives 98.4% highest accuracy.

V. Conclusion

The proposed paper predicts most severe cancer as Brest cancer, which causes many deaths of girls/women's in the world. Hence early detection of disease will save plenty of lives. The proposed system identifies breast cancer by SVM Random forest methods. This paper concentrates on comparison of the 2 SVM and RF methods. By use of Python, SVM classifies the dataset on implementation gives accuracy of 96.68% in training phase. RF has accuracy 98.4% on the average. This research work has compared the Random graph values, which was obtained on 5,10 and 20 folds dataset also the graph shown states the accuracy level of the RF method over the SVM. Researchers and engineers in data processing can have the benefit of this survey. Hence, this research work has concluded by comparing the graph and results obtained that, random forest is way superior to SVM methodology for the detection of carcinoma.

PROFILE OF AUTHORS

Lakshman(lakshman.korra@gmail.com). Received M.Tech Degree in Computer Science from Jawaharlal Nehru Technological University Anantapur (2007) and B.Tech (2005) from JNTU Hyderabad. Currently working as a Scientist"D" ' in NIELIT Aurangabad. Current research interest areas are in, Image Processing.

Siddharth B. Dabhade, he has completed M.Sc. Information Technology (2009), M.Phil. Computer Science (2012), Ph.D. Computer Science (2018). He has published more than 30 research articles in reputed national, international conferences and journals. He is member of various learn bodies like IEEE, ISCA, IAPR, CSTA, AITP, IACSIT, IAENG, IAPR, European Association for Signal Processing, etc. On the basis of his research work he has got 16 distinct awards like Young Scientist Award, Best Paper presentation, excellent paper, etc. Financially he has funded by UGC, DST-SERB ITS, International Association for Pattern Recognition (IAPR), Morpho and Eurasip, therefore, he has visited Italy, Japan, Malaysia countries to represent his research work.

Y S Rode is currently working as a Assistant Professor and Head, Department of Computer Science and IT, Jijamata Mahavidyalaya, Buldhana, MS, India

Karan dabhade, Received M.E. in CSE, presently working as Junior Faculty at NIELIT, Aurangabad

Dr. Sachin N. Deshmukh is currently working as a Professor and Head, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India. He is also working as a Director, IQAC and Director, Bajaj Incubation Center. He has worked as Director, University Network & Information Center (UNIC), Director, Center for Vocational Education and Training, Chief Coordinator of Spoken Tutorial Project of IITB and PET-2014 Coordinator, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India

Dr. Ranjan Maheshwari, (ranjan@rtu.ac.in) Ph.D, Professor at RTU, Kota, India. He is also worked as Director of NIELIT Aurangabad and principal of Government Engineering college Ajmer

REFERENCES

[1] Karthikeyan, S. Venkatakrishnan" carcinoma Classification using SVM Classifier" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019

[2] Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan" Prediction of carcinoma Using Support Vector Machine and K-Nearest. 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka,Bangladesh.

[3] Y.Ireaneus Anna Rejani "Early Detection Of carcinoma Using Svm Classifier" International Journal on technology and Engineering Vol.1(3), 2009,127-130

[4] Simon Tong, Daphne Koller, "Support Vector Machine Active Learning with Applications to Text Classification" Journal of Machine Learning Research (2001)45-66. DOI: 10.1162/153244302760185243

[5] Theodoros Evgeniou, Massimiliano Pontil "Support Vector Machines: Theory and Applications" Conference: Machine Learning and Its Applications, Advanced Lectures January 2001,DOI: 10.1007/3-540-44673-7_12

[6] Issam El-Naqa, Yongyi Yang, Miles N. Wernick, ,Nikolas P. Galatsanos, Robert M. NishikawaA "Support Vector Machine Approach for Detection of Micro calcifications" IEEE Transactions On Medical Imaging, Vol. 21, No. 12, December 2002

[7] M. K. Hasan, M. M. Islam and M. M. A. Hashem, "Mathematical model development to detect carcinoma using multigene genetic programming 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, pp. 574-579, 2016

[8] E.C.Fear, P. M. Meaney, and M. A. Stuchly, "Microwaves for carcinoma detection", IEEE potentials, vol.22, pp.12-18, February-March 2003.

[9] N. Acir, "A support vector machine classifier algorithm supported a perturbation method and its application to ECG beat recognition systems" Expert systems with application big apple, vol.31, pg. 150-158 July2006.

[10] Satish Saini and Ritu Vijay, "Design and Analysis of carcinoma Detection System Using Mammogram Features Extraction", in International Journal of Electronics and computing Engineering, Vol. 3, Number 4, 2015, pp 406-409.

[11] J. Blagojce, K. Ivan, T. Katarina, D. Ivika and L. uzana, "Mammographic Image Classification Using Texture Features", 9th conference for Informatics and data Technology,(2012).

[12] Cuong Nguyen , Yong Wang, Ha Nam Nguyen," Random forest classifier combined with feature selection for carcinoma diagnosis and prognostic" J. bioscience and Engineering, 2013, 6, 551-560 Published Online May 2013

[13] Yingjie Tian , Yong Shi , Xiaohui Liu " Recent Advances on Support Vector Machines Research Technological and economic development of Economy2012 Volume 18(1): 5–33

[14] Vikas Chaurasia , Saurabh Pal , BB Tiwari, "Prediction of benign and malignant carcinoma using data processing techniques " Journal of Algorithms & Computational Technology 2018, Vol. 12(2) 119–126.

[15] Reka Toth,Heiko Schiffmann, Claudia Hube-Magg, Franziska Büscheck, Doris Höflmayer, Sören Weidemann ,"Random forest-based modelling to detectbiomarkers for glandular cancer progression" Toth et al. Clinical Epigenetics (2019) 11:14

[16] Sebastien Jean Mambou, Petra Maresova, Ondrej Krejcar, Ali Selamat, Kamil Kuca "Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model" 2018 Aug 25. DOI: 10.3390/s18092799

[17] Suma, V. "A Novel Information retrieval system for distributed cloud using Hybrid Deep Fuzzy Hashing Algorithm." JITDW 2, no. 03 (2020): 151-160

[18] Vijayakumar, T. "Posed Inverse Problem Rectification Using Novel Deep Convolutional Neural Network." Journal of Innovative Image Processing (JIIP) 2, no. 03 (2020): 121-127.