KDD ON CXR Images of COVID 19 by JPEG Coefficient Filter Technique

S. M. Manimegalai [1], Dr. T. Ramaprabha [2]

[1] Research Scholar, PG and Research Department of Computer Science and Computer Application, Vivekanandha College of Arts and Science for Women (A), Elayampalayam, Tamil Nadu, India.

[2] Professor, PG and Research Department of Computer Science and Computer Application, Vivekanandha College of Arts and Science for Women (A), Elayampalayam, Tamil Nadu, India.

[1] manimegalai352022@gmail.com, [2] ramaradha1971@gmail.com

Article Info
Page Number: 1086-1096
Publication Issue:
Vol. 71 No. 3s (2022)

Article History
Article Received: 22 April 2022
Revised: 10 May 2022
Accepted: 15 June 2022
Publication: 19 July 2022

Abstract

Prevent the spread of virus utilizing computer-based analysis for the goal of fast and reliable identification of corona virus disease (COVID-19). Chest X-ray imaging has many advantages in image processing methods, such as low cost, portability, speed, and ease of use. The impact of many image enhancement techniques are investigated in this research. The impact of many image enhancement techniques are investigated in this research. The impact of many common picture enhancement techniques is investigated in this research. The Random Forest (RF) classifier produces the highest accuracy value, positive predictive value, true positive rate, Area under the receiver operating characteristic curve value, Area under the precision recall curve value which are 84.03% accuracy level, 0.86 of positive predictive value, 0.84 of true positive rate value level, 0.96 of area under the receiver operating characteristic curve value, and 0.91 of area under the precision recall curve value This research work finds that the Random Forest learning model is most recommended model with JPEG Coefficient Filtering technique for CXR image classifications of COVID - 19 Image dataset.

Keywords: JPEG coefficient filter, SMO classifier, IBK classifier, bayes net and tree classifier.

1. Introduction

The global healthcare systems have been overburdened by the Coronavirus Disease 2019 (COVID19) pandemic, which has an exponential infection rate [1,2,3]. The Reverse Transcription Polymerase Chain Reaction (RTPCR) is used to diagnose COVID19, but it has limited accuracy, latency, and sensitivity [4, 5, 6]. Early analysis of an infection improves the likelihood of positive cure for infested patients and also minimizes the risk of an infectious disease like COVID19 spreading in the community.

Radiography image such as chest X-rays (CXR) or computed tomography (CT) are common diagnostic tools for lung disorders such as pneumonia [8, 9], TB [10], and also used to detect COVID19 [1, 11]. One of the benefits of CXR is the ease with which it may be performed utilizing portable X-ray devices, allowing for sooner and more precise COVID 19 diagnosis [10, 11-14]. Artificial intelligence (AI) has discovered that CXRs are capable of identifying COVID19 and less hazardous to the human than CT scans.

The following is how the rest of the article is structured: Segment 2 depicts related works, whereas Segment 3 depicts materials and procedures. The discussions and results are accessible in Segment 4; finally, in section 5, the conclusion is presented.

2. Related Works

COVID -19[1,2,3-13] is one of the most pressing concerns confronting the global health-care system. Many successful methods for diagnosing COVID-19 using chest radiography pictures, such as CT scans and X-rays, have been proposed. A vast number of studies have recently been conducted to identify COVID19 utilizing X ray images and AI models. A small dataset of COVID19 CXR images was developed by several researchers to train machine learning algorithms for automated COVID19 identification [14-23].

For COVID-19 positive detection, Motamed et al. [24] suggested a semi-unsupervised generative model (RANDGAN). The X-ray scans of the chest are shown here. Using 14,100 CXR and an AUC of 0.77, a randomised generative adversarial network (RANDGAN) distinguishes pictures of unidentified classes (COVID19) from recognized and categorized programs (Viral Pneumonia and Normal) without the use of label and preparation data from the unidentified class of image (COVID19).

In a dataset of 1427 X rays, Ioanis et al. [25] found 96.78 percent accurateness for COVID19 from Normal X rays and Bacterial Pneumonia. With a short library of 196 X ray images, Abbas et al. [26] claimed an accurateness of 95.12% for COVID19 organization from COVID19, Regular, and SARS CXRs using their CNN models. Using the ChexPert dataset, Minaee et al. [27] and Irvin et al. [28] obtained a specificity and compassion of 90% and 97 percent, correspondingly. These findings demonstrated the ability of CNN to differentiate COVID 19 from other lung disorders using CXR pictures.

3. Materials and Methods

This section emphases on the materials and methods of this research work. The dataset namely Covid 19 was acquired from the Kaggle repository. The following table represents that the meta data of collected dataset.

Table 1. Overall Information of Dataset

S. No	Class	Size of Image	No of Images	10% (randomly selected) from borrowed set
1	Viral Pneumonia Image	256x256	1345 / 21170	135/ 2117
2	Lung Opacity Image	256x256	6012/ 21170	601/ 2117
3	Normal Image	256x256	10192/ 21170	1019/ 2117
4	Covid19 Image	256x256	3616 / 21170	362 / 2117

4. Methods

The following approaches are applied in this research work. [24-27]

1) Image Collection
2) Image preprocessing
3) An Implement JPEG Coefficient Filter technique
4) Implement the selected learning. i.e.,
    a. BayesNet,
    b. SMO,
    c. IBK,
    d. Bagging,
    e. JRip,
    f. Random Forest
5) Get an efficient outcomes
6) Find a best Model

To produce best outcome, these strategies are implemented in leading open source data mining tool, i.e., Weka.3.9.5. This work considers only 10% of the total dataset and implements 10 fold cross validation for all selected learning techniques.

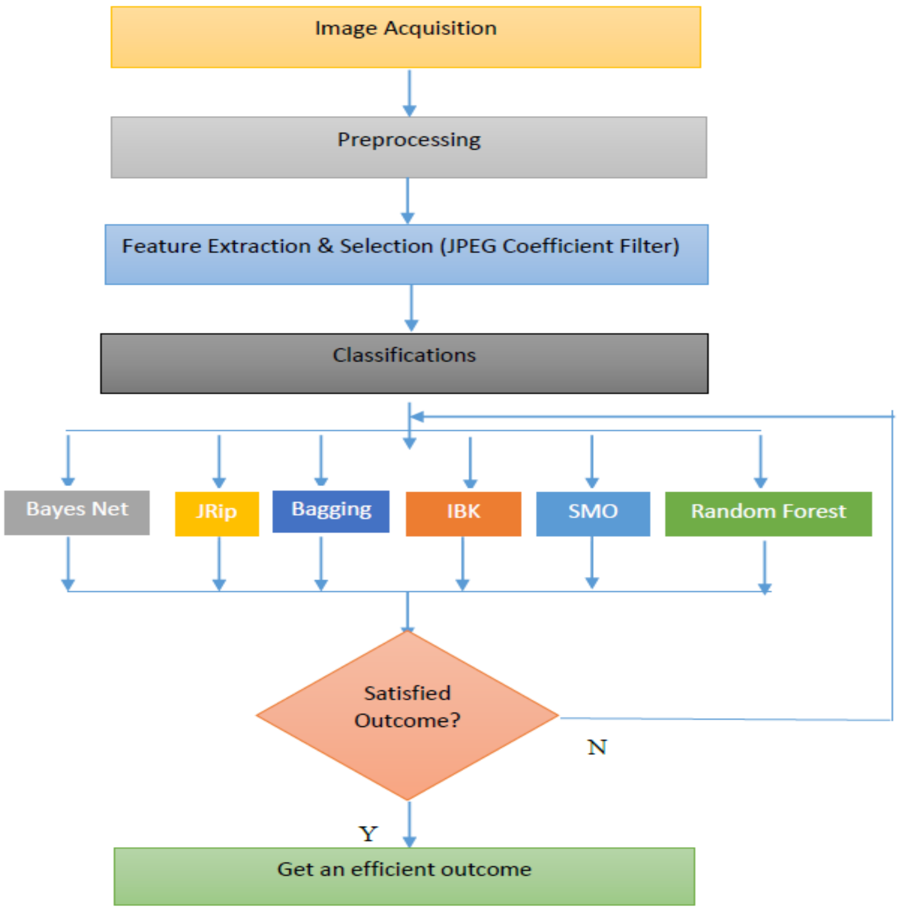

Figure 1; Proposed System Architecture

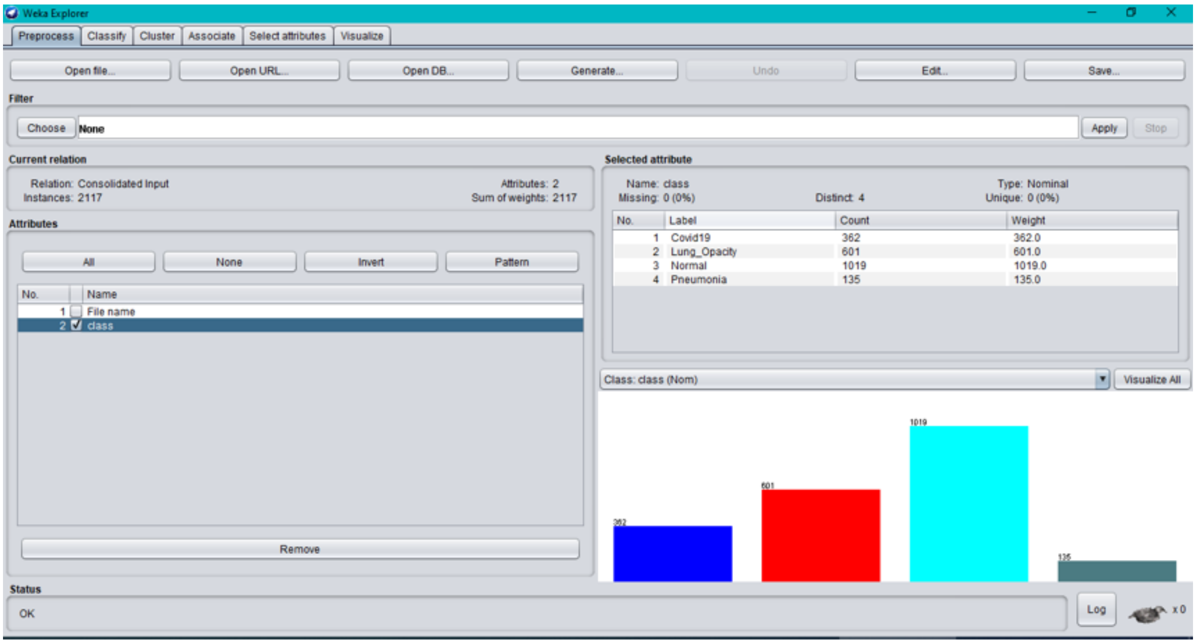

Figure 2; Overview of Multiclass in Weka3.9.5

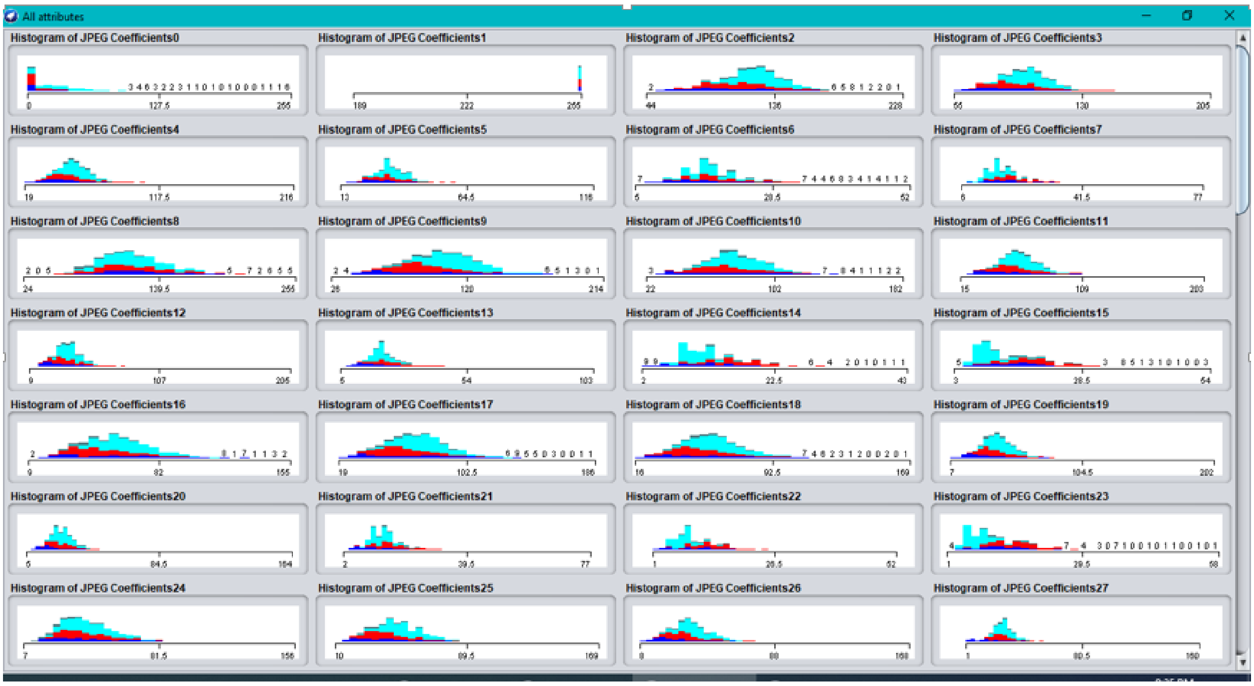

Figure 3; Visualization of Attributes Derived from Dataset by Using JPEG Coefficient Filter

5. IV Results and Discussions

This section presents that the results and interpretations of this research work. By using selected learning models namely, Bayes Net learning model, SMO learning model, Instance Based Classifier, Bagging learning model, Jrip learning model in JPEG Coefficient filter are used to get an efficient outcome of this research work. This work discusses of Accuracy, Positive Predictive value, True Positive Rate, Area under the ROC, Area under the PRC, and time duration for constructing the model of various learning techniques which as mentioned in figure 1, 2 and 3.

Table 3. Performance of Various Classifiers on Covid-19 Dataset by JPEG Coefficient Filter

S. No	**Classifier**	**Accuracy**	**Positive Predictive value**	**True Positive Rate**	**AUCROC**	**AUCPRC**	**Time Taken to build**
1	Bayes Net	77.18%	0.78	0.77	0.92	0.81	0.19
2	Sequential Minimal Optimization	83.65%	0.84	**0.84**	0.91	0.77	1.09
3	IBK	82.48%	0.82	0.83	0.87	0.74	**0.00**
4	Bagging	82.57%	0.83	0.83	0.95	0.88	0.23
5	Jrip	79.68%	0.79	0.80	0.88	0.77	4.31
6	Random Forest	**84.03%**	**0.86**	**0.84**	**0.96**	**0.91**	1.44

The above table shows that Bayes Net learning model of Bayes category produces 77.18% of accuracy value, 0.78 of positive predictive value level, 0.77 of true positive rate value, 0.92 of AUC-ROC value, 0.81% of AUC-PR and it is consuming 0.19 seconds to make a model.

The Sequential Minimal Optimization learning model of Function category produces 83.65% of accurateness level, 0.84 of positive predictive value, 0.84 of true positive rate value, 0.91 of area under the ROC value, 0.77 of area under the PRC and it is consuming 1.09 seconds to construct a model.

The K Nearest Neighborhood learning model (IBK) of Lazy category produces 82.48% of accuracy value, 0.82 of positive predictive value, 0.83 of true positive rate value, 0.87 of area under the ROC, 0.74 of area under the PRC and it is consuming zero seconds to create a model.

The Bagging learning model of Meta category produces 82.57% of accuracy value, 0.83 of positive predictive value, 0.83 of true positive rate value, 0.95 of area under the ROC value, 0.88 of area under the PRC value and it is consuming 0.23 seconds to generate a model.

The Jrip learning model of Rules category produces 79.68% of accuracy value, 0.79 of positive predictive value, 0.80 of true positive value, 0.88 of area under the ROC value, 0.77 of area under the PRC value and it is consuming 4.31 seconds to produce a model.

The Random Forest learning model of Trees category produces 84.03% of accuracy level, 0.86 of positive predictive value, 0.84 of true positive rate value, 0.96 of area under the ROC value, 0.91 of area under the PRC value and it is consuming 1.44 seconds to form a model.

ISSN: 2094-0343
2326-9865

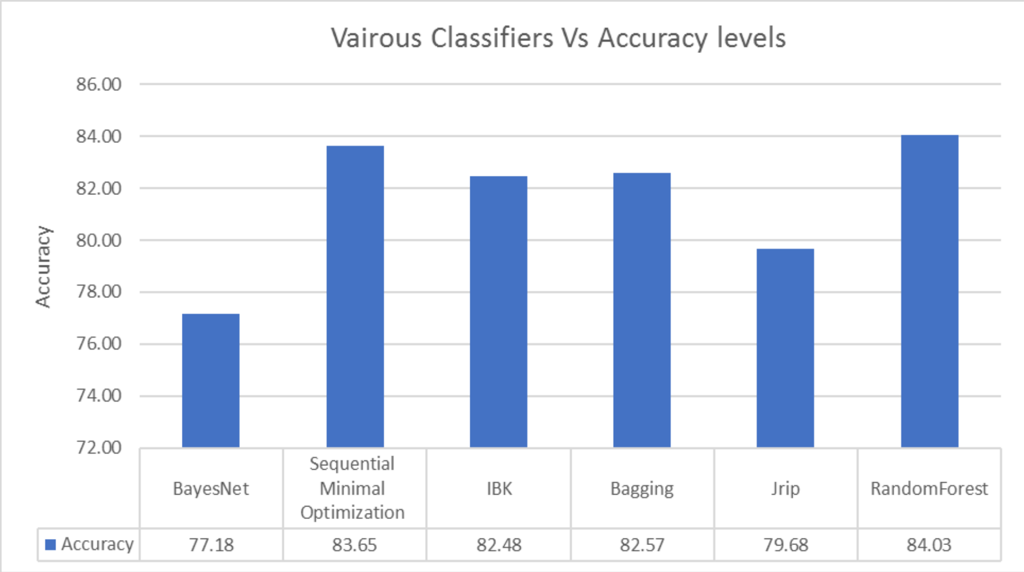

Figure 4; Performance of Accuracy Level of Various Classifiers

The above Figure 4 clearly demonstrates that the Bayes Net Classifier of Bayes classification has the lowest accuracy level of 77.18%, while Random Forest (RF) of Tree classification has the highest accuracy level of 84.03%. The SMO of function classification has an accuracy level of 83.65%, the Instance Based Classifier of the Lazy group has an accuracy level of 82.48%, the Bagging model from the meta group classifier has an accuracy level of 82.57%, and the Jrip from the rules classifier has an accuracy value of 79.68%.

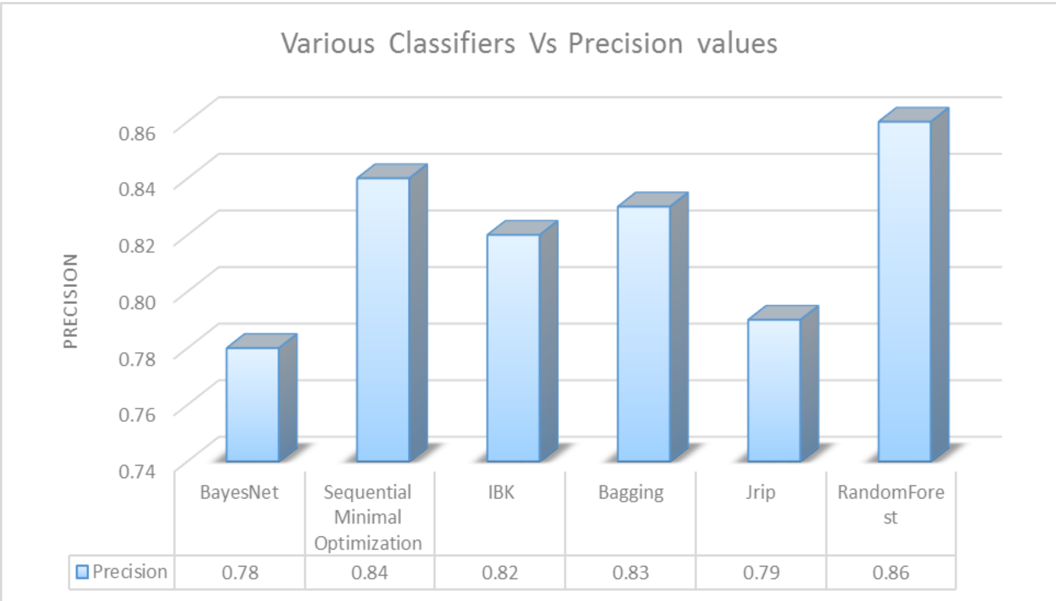

Figure 5; Performance of Precision (Positive Predictive Value) of various Classifiers

The above diagram Figure 5 shows that the Bayes Net learning model is yielding 0.78 positive predictive value (PPV) level this is the least PPV and Random Forest learning model is yielding 0.86 of positive predictive value. This the highest value of PPV compare with other learning models.

The SMO learning model is cropping 0.84 of PPV level, the Bagging learning model is having 0.83 of PPV, the Instance Based Classifier learning model is holding 0.82 of PPV level and the Jrip of rules classification is consuming 0.79 of PPV.

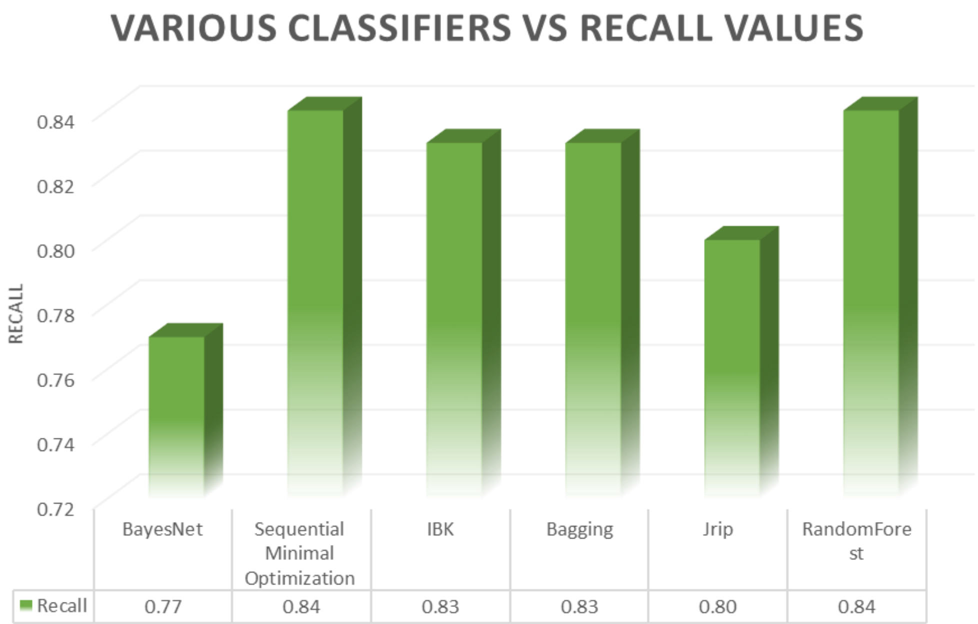

Figure 6; Performance of Recall (True Positive Rate) Value of various Classifiers

The above figure 6 shows that the Bayes Net algorithm of Bayes Classification produces lowest true positive rate value which is 0.77 and the highest true positive rate value is 84% of recall value which is shown by Sequential Minimal Optimization learning model and Random Forest learning model. The Instance Based classification learning model and Bagging learning models are producing same true positive rate value which is 0.83, The Jrip learning model is making 80% of recall value.

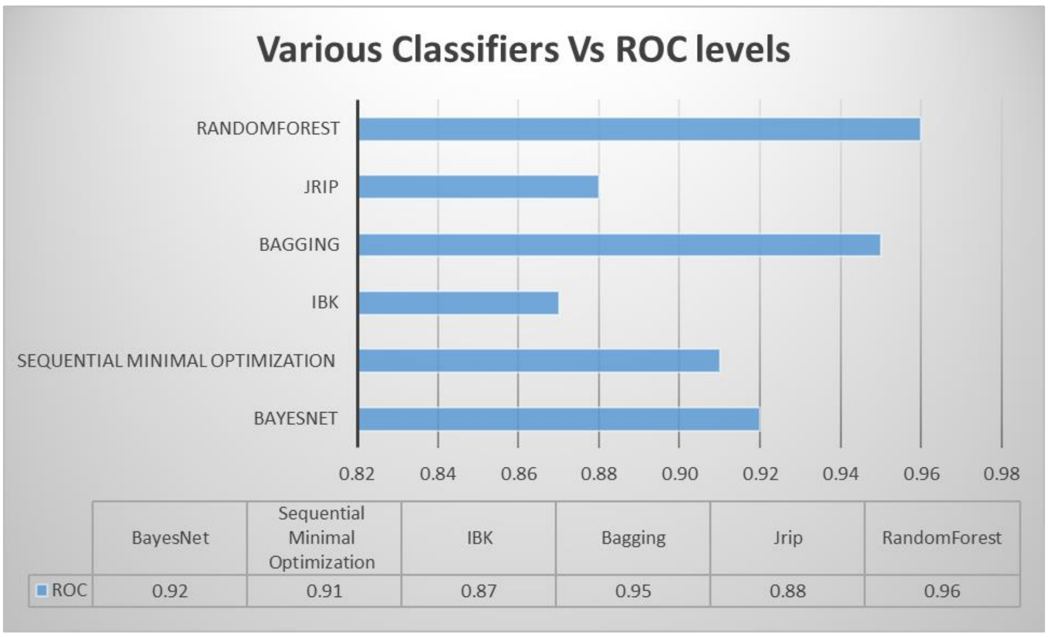

Figure 7; Performance of AUCROC (ROC) Value of Selected Classifiers

The above figure 7 shows that the Random Forest learning model is cropping 0.96 of AUCROC which is maximum AUCROC value level compare with other models and least AUC ROC is 0.87 which is made by Instance based classification learning model.

The Sequential Minimal Optimization of purpose classification is having 0.91 of AUC-ROC value level, the Bagging learning model of collaborative classification is having 0.95 of AUC-ROC value level, The Bayes Net algorithm of Bayes classification is holding 0.92 of AUC-ROC value level and the Jrip learning model is gathering 0.88 of AUCROC value.

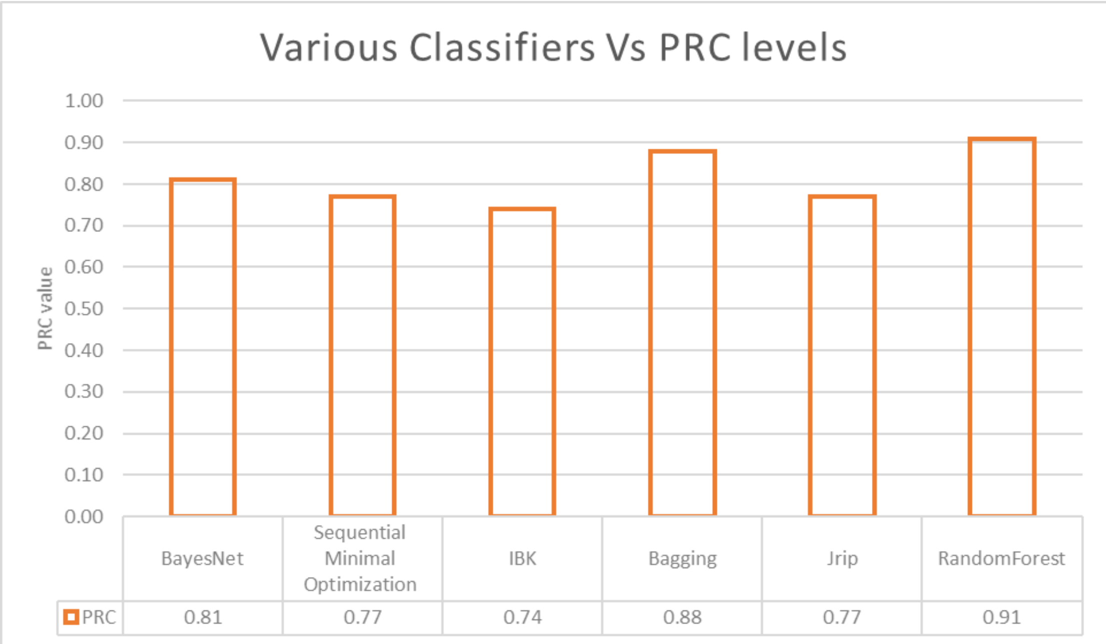

Figure 8; Performance of AUCPRC (PRC) Value of various Classifiers

The above graph 8 shows that the lowest area under the precision recall value is 0.74 which is owned by IBK learning model of lazy category and the highest PRC area under curve value is 0.91 of PRC area under curve which is given by Random Forest classifier.

The Bayes Net classifier is producing 0.81 of AUCPRC, The SMO of function category classification and Jrip of rules category classification are producing same PRC area under curve value which is 0.77 of PRCAUC value, and the Bagging algorithm of meta classification category is producing 0.88 of AUCPRC value.

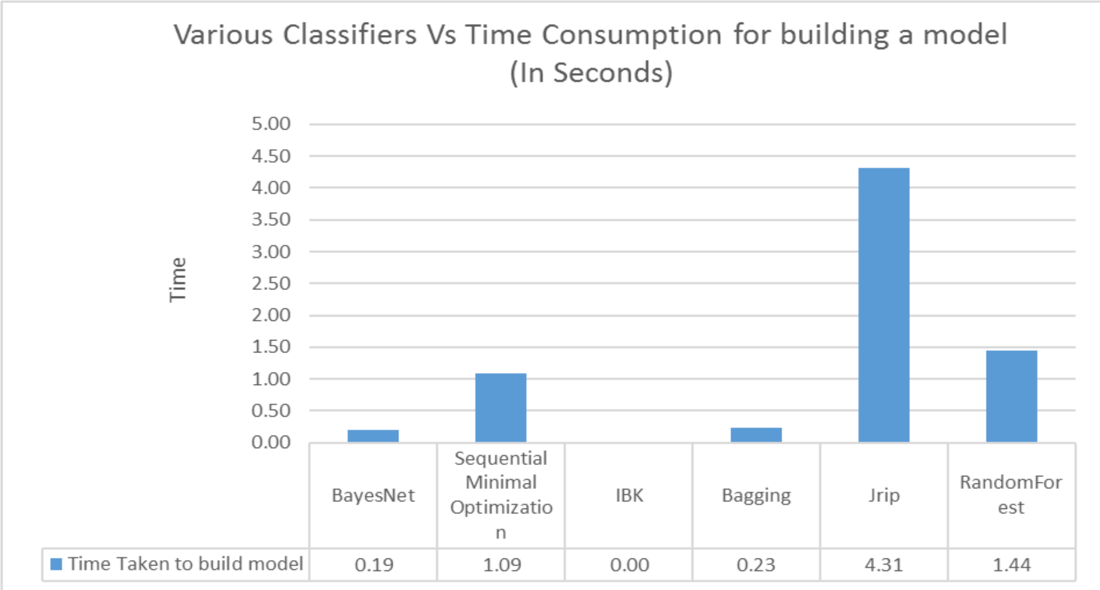

Figure 9; Performance of time Consumption to Build Model of various Classifiers

The above diagram figure 9 shows that the IBK classification algorithm of lazy category takes a zero seconds to build a model which is less time ingesting for building a model compare with other models and Jrip classifier takes more time ingesting to build a model which is 4.31 second.

Random Forest classification of trees classification category takes 1.44 seconds to build a model, The SMO algorithm of function classification category takes 1.09 seconds to build a model, The Bagging algorithm of meta classification category takes 0.23 seconds to build a model and the Bayes Net algorithm of Bayes classification category takes 0.19 seconds to build a model.

6. Conclusion

This study discovered that the Bayes Net Classifier has the lowest accuracy level of 77.18 percent, while the Random Forest classifier has the highest accuracy level of 84.03%. The Bayes Net classifier produces the lowest positive predictive value level of 0.78, while the Random Forest of Trees category produces the highest positive predictive value of 0.86. The Bayes Net classifier has the lowest true positive rate value of 0.77, while the SMO learning and the Random Forest learning have the best true positive rate value of 0.84. The Random Forest classifier produces the highest ROC value of 0.96, while the IBK classifier produces the lowest ROC value of 0.87. The IBK classifier has the lowest PRC value of 0.74, while the Random Forest classifier has the greatest. The IBK classifier takes zero seconds to build a models, which is less time than other model, and the Jrip classifier takes 4.31 seconds to build a models, which is more time than other models. This research work evaluate various performances of various classifiers and recommends that the Random Forest classifier produces an optimal solutions compare with other models.

References

1. M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M.T. Islam, "Can AI help in screening Viral and COVID-19 pneumonia?" IEEE Access, Vol. 8, 2020, pp. 132665 - 132676.
2. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2020. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images. arXiv preprint arXiv:2012.02238.
3. Alshazly H, Linse C, Abdalla M, Barth E, Martinetz T. 2021. COVID-Nets: deep CNN architectures for detecting COVID-19 using chest CT scans. PeerJ Comput. Sci. 7:e655 http://doi.org/10.7717/peerj-cs.655
4. S.A. Harmon, T.H. Sanford, S. Xu, E.B. Turkbey, H. Roth, Z. Xu, et al., "Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets", Nature Communications, vol. 11(1), p. 4080(2020/08/14 2020)
5. Y. Yang, M. Yang, C. Shen, F. Wang, J. Yuan, J. Li, et al., "Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. medRxiv 2020.02. 11.20021493 [Preprint]. 17 February 2020," Preprint] doi: 0.1101/2020.02, vol. 11)
6. M. Yu, D. Xu, L. Lan, M. Tu, R. Liao, S. Cai, et al., "Thin-section Chest CT Imaging of Coronavirus Disease 2019 Pneumonia: Comparison Between Patients with Mild and Severe Disease," Radiology: Cardiothoracic Imaging, vol. 2(2), p. e200126(2020)
7. T. Rahman, M.E. Chowdhury, A. Khandakar, K.R. Islam, K.F. Islam, Z.B. Mahbub, et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray," Applied Sciences, vol. 10(9), p. 3233(2020)
8. T. Rahman, A. Khandakar, M.A. Kadir, K.R. Islam, K.F. Islam, R. Mazhar, et al., "Reliable Tuberculosis Detection using Chest X- ray with Deep Learning, Segmentation and Visualization", IEEE Access, vol. 8 pp. 191586-191601(2020)
9. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, et al., "Coronavirus: Comparing COVID-19, SARS and MERS in the eyes of AI," arXiv preprint arXiv:2005.11524, 2020)
10. M.E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, et al., "Can AI help in screening viral and COVID-19 pneumonia?," IEEE Access, vol. 8 pp. 132665-132676(2020)
11. Pawan Kumar Tiwari, P. S. . (2022). Numerical Simulation of Optimized Placement of Distibuted Generators in Standard Radial Distribution System Using Improved Computations. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 10–17. https://doi.org/10.17762/ijrmee.v9i5.369
12. T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," Radiology, p. 200642(2020)

13. Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR", Radiology, p. 200432(2020)
14. M.E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, et al., "Can AI help in screening viral and COVID-19 pneumonia?," arXiv preprint arXiv:2003.13145, 2020)
15. https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711
16. https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png
17. https://sirm.org/category/senza-categoria/covid-19/
18. https://eurorad.org
19. https://github.com/ieee8023/covid-chestxray-dataset
20. https://figshare.com/articles/COVID-19_Chest_X-Ray_Image_Repository/12580328
21. https://github.com/armiro/COVID-CXNet
22. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
23. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
24. WEKA: http://www.cs.waikato.ac.nz
25. S. Motamed, P. Rogalla, and F. Khalvati, ″RANDGAN: Randomized Generative Adversarial Network for Detection of COVID-19 in Chest X-ray, arXiv preprint arXiv:2010.06418, 2020)
26. I.D. Apostolopoulos and T.A. Mpesiana, ″Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks", Physical and Engineering Sciences in Medicine, p. 1(2020)
27. Abbas, M.M. Abdelsamea, and M.M. Gaber, ″Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," arXiv preprint arXiv:2003.13815, 2020)
28. S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G.J. Soufi, ″Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," arXiv preprint arXiv:2004.09363, 2020)
29. J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, et al., "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 590-597.
30. Ayyappan.G et al. "Heart Disease Data Set Classifications: Comparisons of Correlation Co Efficient by Applying Various Functions," Indian Journal of Computer Science and Engineering (IJCSE), Volume No. 9 Issue No. 5 Oct-Nov 2018, Page No: 135-140, e-ISSN: 0976-5166, p-ISSN: 2231-3850.
31. S.M. Manimegalai et al. "Image Classification on Covid 19 CXR Images Using Auto Color Correlogram Filter", Indian Journal of Computer Science and Engineering (IJCSE), EISSN: 0976-5166 P-ISSN: 2231-3850 Vol-12, No-5, and Sep-Oct 2021.
32. Vanitha, D. D. . (2022). Comparative Analysis of Power switches MOFET and IGBT Used in Power Applications. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 01–09. https://doi.org/10.17762/ijrmee.v9i5.368