International Conference on
Innovations in Computational Intelligence and Computer Vision (ICICV-2020)
January 17-19, 2020

Prognosis of Breast Cancer by implementing Machine Learning Algorithms using Modified Bootstrap Aggregating

Peeyush Kumar, Ayushe Gangal & Sunita Kumari
GGSIPU

CONTENTS

INTRODUCTION & PURPOSE

- Breast cancer is the second most perilous type of cancer after lung cancer.
- Early diagnosis is conducive to ameliorate the breast cancer outcomes and survivability.
- A major challenge faced by the doctors and practitioners worldwide is the identification and diagnosis of the cancerous cells.

INTRODUCTION & PURPOSE (contd.)

- The implementation of artificial intelligence has made the prognosis of maladies possible, by using pattern recognition as its base concept.

- A hybrid machine learning model is proposed, which makes use of the ability of decision trees to bifurcate the data on the basis of certain features, followed by the application of machine learning algorithms on those resulting datasets.

- The proposed model is created using Deep Learning and Support Vector Machine algorithms, on the data, pre-processed using decision tree algorithm.

- The Wisconsin Breast Cancer dataset from the UCI machine learning repository is used for this study.

EARLIER WORKS

- Dana Bazazeh et al. did a comparative study with three machine learning algorithms, namely, Support Vector Machine, Random Forest and Bayesian Network for the prediction of breast cancer.

- U.K. Kumar et al. classified breast cancer as malignant or benign Naive Bayes, SVM and J48, a type of decision tree.

- Dongdong Sun et al. devised a new machine learning algorithm called as DSVM, created by integrating Deep Learning (DL) and Support Vector Machine (SVM). The database used was obtained from broad GDAC firehose.

EARLIER WORKS (contd.)

- Puneet Yadav et al. used Decision Tree and Support Vector Machine for the classification of breast cancer, as malignant or benign.

- Autsuo Higa classified breast cancer using Decision Tree and Artificial Neural Network (ANN).

- A. Suresh et al. proposed a hybridized model using neural network and decision tree. The main objective was to deal with the misclassified values. Radial Basis Function Network and Decision Tree were used.

EARLIER WORKS (contd.)

- Ch. Shravya et al. classified breast cancer as malignant or benign, by using Logistic Regression, Support Vector Machine (SVM) and K-Nearest Neighbor. Dimensionality Reduction is applied to reduce the independent variables to a set of principal variables.

PROPOSED METHODOLOGY

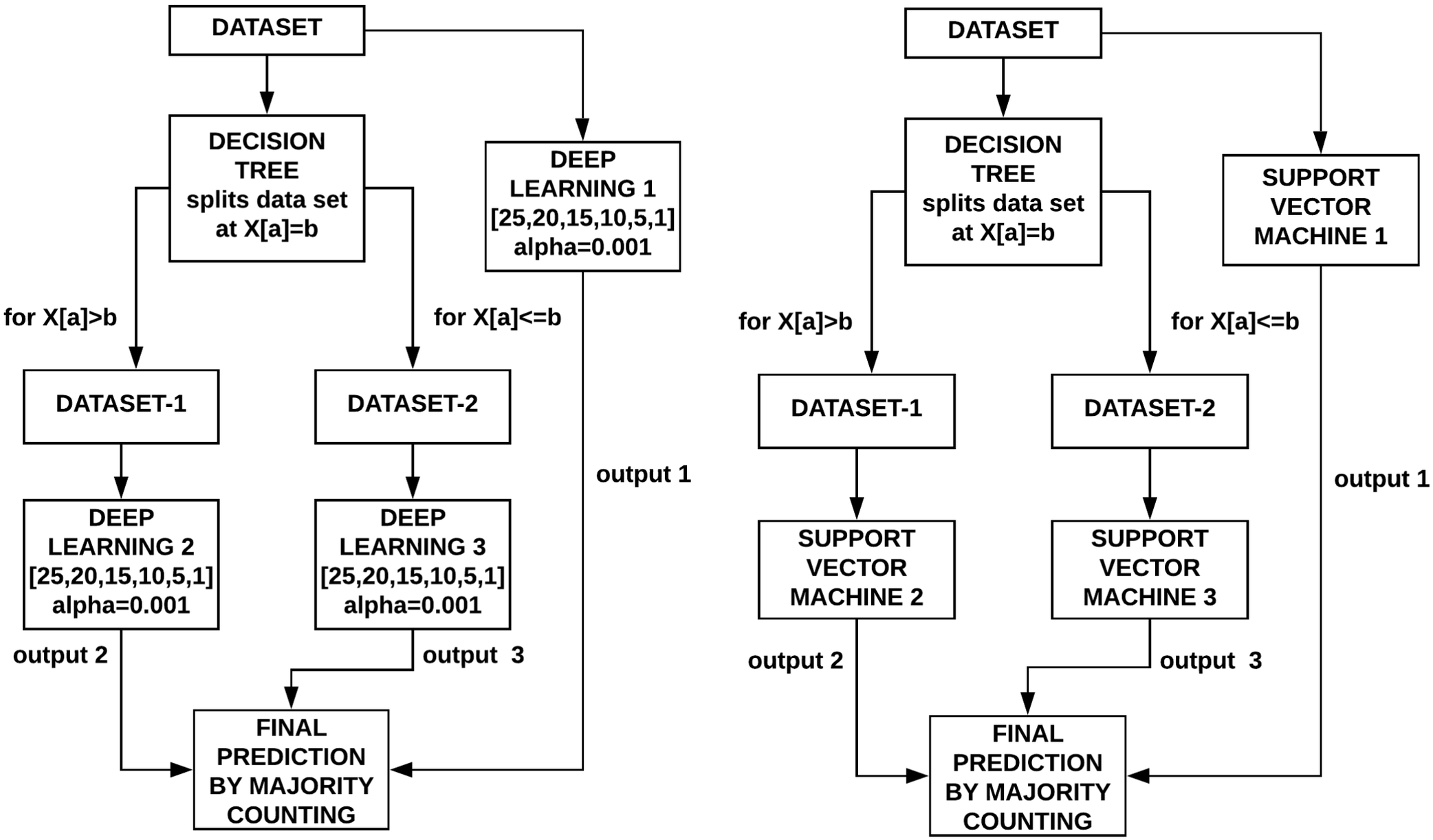


Fig 1: Flowcharts depicting the proposed methodology using Deep Learning and Support Vector Machine.

PROPOSED METHODOLOGY (contd.)

- The proposed model uses a modified way of bootstrap aggregating, which involves randomly picking training examples unlike done in standard bootstrap aggregating.

- A Decision Tree is built by partitioning the data set at points where the majority number of two classes can be separated.

- This bifurcation property of decision tree plays an indispensable role in the proposed model.

ALGORITHM FOR DATA DIVISION

- A dataset L consists of data $\{(y_n , x_n), n = 1….N\}$, where the y's are class labels.
- Using this dataset, a predictor $\varphi(x,L)$ can be formed— if the input is x, we predict y by $\varphi(x,L)$.
- Using these extracted values the dataset is sliced at X[22]=106.1 and X[20]=16.795, for proposed model with SVM and Deep Learning respectively. This results in three datasets as follows:

$$L^B \epsilon \{ L^K \} \quad (1)$$

where, k= 0, 1, 2

Now we get,

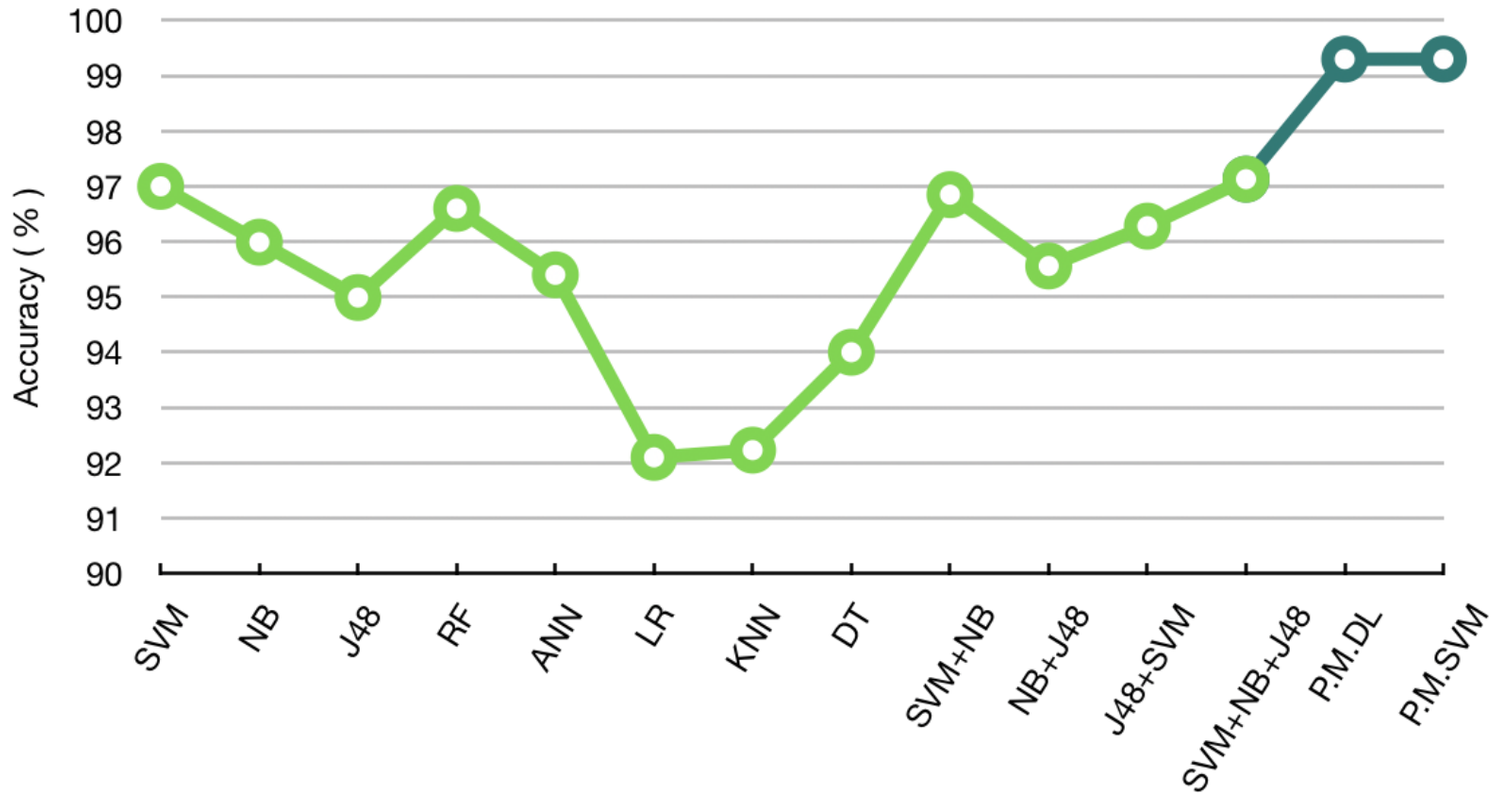
L^1= 70% of the dataset L and rest 30% is a test set
L^2= is formed by randomly removing the training examples with $X[a^0] > b^0$
L^3= is formed by randomly removing the training examples with $X[a^1] <= b^1$

where a^i = i^{th} element of a = [22,20], and b^i = i^{th} element of b = [106.1, 16.795] respectively.

RESULTS AND DISCUSSIONS

- The proposed model achieved 99.3% accuracy for both Deep Learning and Support Vector Machine.

- The graph juxtaposes the accuracy values obtained by the algorithms used previously by the researchers, with the accuracy of the proposed model.



Graph 1. Graph comparing the accuracy values of various algorithms used in previous works, with the accuracy of the Proposed Model.

CONCLUSION, LIMITATIONS AND FUTURE WORKS

- A novel hybrid approach for the prognosis of breast cancer, by modifying the standard bootstrap aggregating is presented.

- The algorithm works by dividing the data before the application of the machine learning algorithms. Thus, obtaining better results in terms of accuracy than any other algorithm.

- The results obtained prove that the proposed method performs better than all the standard bagging models.

- The future expects to use the proposed model for the diagnosis of other types of cancers and other diseases as well. The work done can be used for the real-world of the medical field in clinical practices of the practitioners.

REFERENCES

1. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/, last accessed 2019/09/08.
2. https://www.mayoclinic.org/diseases-conditions/breast-cancer/diagnosis-treatment/drc-20352475, last accessed on 2019/09/24.
3. https://www.webmd.com/breast-cancer/guide/breast-cancer-symptoms-and-types, last accessed on 2019/09/25.
4. Al-hadidi, M.,: Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm. In: 9th International Conference on Developments in eSystems Engineering (2016)
5. Bazazeh, D., et al.,: Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis. In : International Journal of Intelligent Systems and Applications in Engineering (2016)
6. Higa, A.,: Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms. In: International Journal of Computer Applications Technology and Research (2018)
7. Suresh, A., et al.,: Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers. Soft Computing. Springer Nature (2019)
8. Shravya. C., et al.,: Prediction of Breast Cancer Using Supervised Machine Learning Techniques. In: International Journal of Innovative Technology and Exploring Engineering (IJITEE) (2019)
9. Wisconsin Breast Cancer dataset, available at https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+ (Diagnostic), last accessed on 2019/09/24.
10. Yang, Y., et al.,: The Research of the Fast Svm Classifier Method. In: 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) pp. 121-124. (2015)
11. Pagariya, R., et al.,: Review Paper on Artificial Neural Networks. International Journal of Advanced Research in Computer Science, vol. 4, (2013)

REFERENCES

12. Müller, A., Guido, S.,: Introduction to Machine Learning with Python. O'Reilly Publication. Sebastopol, California, USA (2016)

13. Chollet, F.,: Deep Learning with Python, Manning Publications. Shelter Island, New York, USA (2018)

14. Hastie, T., Tibshirani, R., Friedman, J.,: The Elements of Statistical Learning. Springer series in statistics. Berlin, Germany (2009)

15. Ng, A.,: Machine Learning. course on coursera, available at https://www.coursera.org/learn/machine-learning

16. Patel, B., et al.,: A Survey on Decision Tree Algorithm For Classification. In: International Journal of Engineering Development And Research (IJEDR) (2014)

17. Denil, M., et al.,: Narrowing the Gap: Random Forests In Theory and In Practice. In: Proceedings of Machine Learning Research (PMLR) (2014)

18. Breiman, L.,: Bagging Predictors. In: Kluwer Academic Publishers-Plenum Publishers and Springer Machine Learning Series (1996)

19. http://gdac.broadinstitute.org/, last accessed on 2019/09/24

20. Kumar, U., et al.,: Prediction of Breast cancer using voting classifier technique. In: Proceedings of the IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials. Chennai, India (2017)

21. Sun, D., et al.,: Prognosis Prediction of Human Breast Cancer by integrating Deep Neural Network and Support Vector Machine. In: 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK (2016)

22. Yadav, P., et al.,: Diagnosis of Breast Cancer using Decision Tree Models and SVM. In: International Research Journal of Engineering and Technology (IRJET) (2018)