CLASSIFICATION OF BREAST CANCER DISEASE USING BAGGING FUZZY-ID3 ALGORITHM BASED ON FUZZYDBD

NUR FARAHAINA BINTI IDRIS

MASTER OF SCIENCE

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science.

(Supervisor's Signature)

Full Name : MOHD ARFIAN BIN ISMAIL

Position : SENIOR LECTURER

Date : 8 FEBRUARY 2022

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

farahaina

(Student's Signature)

Full Name : NUR FARAHAINA BINTI IDRIS

ID Number : MCP20001

Date : 8 FEBRUARY 2022

CLASSIFICATION OF BREAST CANCER DISEASE USING BAGGING FUZZY-ID3 ALGORITHM BASED ON FUZZYDBD

NUR FARAHAINA BINTI IDRIS

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Science

Faculty of Computing

UNIVERSITI MALAYSIA PAHANG

FEBRUARY 2022

ACKNOWLEDGEMENTS

Thanks to my parents, Idris bin Ismail and Rahinah Wan Hassan, for their endless support throughout my journey to finish my master programme. I want to thank my family members, especially my sister, Nur Farida, for always staying with me and giving me advice. I also want to thank my friends, Aina Umairah, Amirul Ashraf and Syafiqah, who always listen to me when I have a hard time. Last but not least, I would like to thank my supervisor, Dr Mohd Arfian Ismail, that gave much insightful feedback and guided work to a higher level. Thanks for all the endless support. This study was supported by Fundamental Research Grant Scheme (FRGS) with Vot No. FRGS/1/2018/ICT02/UMP/02/2: RDU190113 from Ministry of Education and managed by Universiti Malaysia Pahang.

ABSTRAK

Klasifikasi adalah teknik perlombongan data yang digunakan untuk mengklasifikasikan pelbagai jenis data mengikut kriteria tertentu. Salah satu kaedah pembelajaran mesin yang paling baik untuk menangani masalah klasifikasi adalah pokok keputusan. Terdapat pelbagai algoritma pokok keputusan, tetapi yang paling biasa digunakan adalah Iterative Dichotomiser 3 (ID3), CART, dan C4.5. ID3 mempunyai kelebihan paling banyak di antara ketiga algoritma, terutamanya dalam masa pemprosesan, kerana ia membina pokok terpantas dengan kedalaman pendek. Namun, walaupun kelaziman pokok keputusan dalam mengendalikan masalah klasifikasi, ia mengalami masalah seperti varians yang tinggi dan overfitting yang membawa kepada generalisasi yang lemah. Gabungan kaedah kabur dan algoritma ID3 menguruskan data dengan lebih cekap kerana menggabungkan kedua-dua kelebihan kaedah kabur dan pokok keputusan. Untuk teknik yang dicadangkan dari algoritma FID3-DBD, atribut berterusan dan diskrit (bilangan bulat) akan ditentukan dalam nilai linguistik set kabur, dan kaedah FUZZYDBD digunakan untuk mengatur parameter set kabur. Penggantian dengan label linguistik set kabur dengan keserasian nilai input tertinggi juga telah dilakukan sebelum aruhan pokok berlaku. Teknik yang dicadangkan menyelesaikan batasan algoritma ID3 tradisional yang tidak dapat mengklasifikasikan atribut bernilai berterusan dan, pada masa yang sama, meningkatkan ketepatan klasifikasi. Kaedah bagging kemudiannya diterapkan pada algoritma FID3-DBD untuk mengatasi masalah overfitting dan varians yang tinggi pada pokok keputusan. Empat set data kanser payudara yang akan digunakan untuk menilai ketepatan klasifikasi: set data Kanser Payudara Wisconsin (Asal), set data WDBC (Diagnostik), set data Kanser Payudara Coimbra dan set data Mammographic Mass. Semua set data tersebut diperoleh dari repositori pembelajaran mesin UCI. Objektif kajian ini adalah untuk menyelesaikan batasan algoritma ID3 tradisional yang tidak dapat mengklasifikasikan data berterusan dengan baik dan juga mengatasi masalah varians tinggi dan overfitting. Metodologi penyelidikan ini terdiri daripada empat langkah asas iaitu tinjauan literatur, pengumpulan data, pelaksanaan eksperimen dan penulisan laporan. Algoritma FID3-DBD memperoleh ketepatan klasifikasi 94.362% untuk set data Kanser Payudara Wisconsin (Asal), 94.358% untuk set data WDBC (Diagnostik), 81.119% untuk set data Mammographic Mass dan 64.224% untuk set data Coimbra. Algoritma BFID3-DBD memperoleh ketepatan klasifikasi 96.003% untuk set data Kanser Payudara Wisconsin (Asal), 95.273% untuk set data WDBC (Diagnostic), 81.590% untuk set data Mammographic Mass dan 68.966% untuk set data Coimbra. Kajian ini mengesahkan bahawa algoritma FID3-DBD mampu mengklasifikasikan data berterusan dan algoritma BFID3-DBD mengatasi overfitting, mengurangkan varians tinggi, dan meningkatkan ketepatan klasifikasi data ujian.

ABSTRACT

Classification is a data mining technique used to classify varied data types according to a specific criterion. One of the most powerful machine learning methods to handle classification problems is the decision tree. There are various decision tree algorithms, but the most commonly used are Iterative Dichotomiser 3 (ID3), CART, and C4.5. ID3 has the most advantages among the three algorithms, especially in processing time, as it builds the fastest tree with short depth. However, despite the decision tree's commonness in handling classification problems, it suffers problems like high variance and overfitting, leading to poor generalisation. The combination of fuzzy and ID3 algorithm manages the data more efficiently as it combines both the advantages of fuzzy and decision tree. For the proposed technique of the FID3-DBD algorithm, the continuous and discrete (integer) attributes would be defined in the linguistic values of the fuzzy sets, and the FUZZYDBD method is being used to set up the fuzzy sets' parameters. Replacement with the linguistic labels of fuzzy sets with the highest compatibility of input values has also been done before the tree induction occurs. The proposed technique solves the limitation of the classic ID3 algorithm that cannot classify the continuous-valued attributes and, at the same time, increase the classification accuracy. The bagging method was then applied to the FID3-DBD algorithm to overcome overfitting problems and high variance in decision trees. Four breast cancer datasets were used to evaluate the classification accuracy: Wisconsin Breast Cancer (Original) dataset, WDBC (Diagnostic) dataset, Breast Cancer Coimbra dataset, and Mammographic Mass dataset. All those datasets were acquired from the UCI machine learning repository. This study aims to solve the limitation of the classic ID3 algorithm that is unable to classify continuous data well and overcome the high variance and overfitting issues. This research methodology consists of four fundamental steps: literature review, data collection, experiment implementation, and report writing. The FID3-DBD algorithm acquired the classification accuracy of 94.362% for the Wisconsin Breast Cancer (Original) dataset, 94.358% for the WDBC (Diagnostic) dataset, 81.119% for the Mammographic Mass dataset and 64.224% for the Coimbra dataset. The BFID3-DBD algorithm obtained the classification accuracy of 96.003% for the Wisconsin Breast Cancer (Original) dataset, 95.273% for the WDBC (Diagnostic) dataset, 81.590% for the Mammographic Mass dataset and 68.966% for the Coimbra dataset. The study verified that the FID3-DBD algorithm could classify the continuous data, and the BFID3-DBD algorithm overcame the overfitting issue, reduced high variance, and increased test data classification accuracy.

TABLE OF CONTENT

REFERENCES

Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019). Model ensembles of artificial neural networks and support vector regression for improved accuracy in the prediction of vegetation conditions and droughts in four northern Kenya counties. *ISPRS International Journal of Geo-Information*, *8*(12). https://doi.org/10.3390/ijgi8120562

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

Al-Ibrahim, A. (2011). Discretization of Continuous Attributes in Supervised Learning Algorithms. *The Research Bulletin of Jordan ACM*, *Vol II (IV*(Iv), 1158.

Al-Salihy, N. K., & Ibrikci, T. (2017). Classifying breast cancer by using decision tree algorithms. *Proceedings of the 6th International Conference on Software and Computer Applications - ICSCA '17*, 144–148. https://doi.org/10.1145/3056662.3056716

Ali, S. K., & Mutlag, W. K. (2018). Early detection for breast cancer by using fuzzy logic. *Journal of Theoretical and Applied Information Technology*, *96*(17), 5717–5728.

Aliev, R. A., Pedrycz, W., Guirimov, B. G., Aliev, R. R., Ilhan, U., Babagil, M., & Mammadli, S. (2011). Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization. *Information Sciences*, *181*(9), 1591–1608. https://doi.org/10.1016/j.ins.2010.12.014

Alshraideh, M. (2017). Automated Detection of Breast Cancer Using Artificial Neural Networks and Fuzzy Logic. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *June*.

Angayarkanni, A. S. P., & Kamal, N. B. (2012). MRI mammogram image classification using ID3 algorithm. *IET Conference Publications*, *2012*(600 CP), 1–5. https://doi.org/10.1049/cp.2012.0464

Anyanwu, M. N., & Shiva, S. G. (2009). Comparative Analysis of Serial Decision Tree Classification Algorithms. *International Journal of Computer Science and Security*, *3*(3), 230–240. http://www.cscjournals.org/csc/manuscriptinfo.php?ManuscriptCode=72.73.66.82.82.44.55.49.99

Assiri, A. S., Nazir, S., & Velastin, S. A. (2020). Breast Tumor Classification Using an Ensemble Machine Learning Method. *Journal of Imaging*, *6*(6). https://doi.org/10.3390/JIMAGING6060039

Austria, Y. D., Goh, M. L., Sta. Maria Jr., L., Lalata, J.-A., Goh, J. E., & Vicente, H. (2019). Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset. *International Journal of Simulation- Systems, Science*

and Technology- IJSSST, *20*, 1–8. https://doi.org/10.5013/ijssst.a.20.s2.23

Avdeenko, T., Makarova, E., & Begenova, S. (2018). Combination of case-based reasoning and data mining through integration with the domain ontology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10943 LNCS*, 159–167. https://doi.org/10.1007/978-3-319-93803-5_15

Badiang, R. O., Gerardo, B. D., & Medina, R. P. (2019). Relocating local outliers produced by K-means and K-medoids using local outlier rectifier V.2.0. *2019 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2019*, 89–94. https://doi.org/10.1109/ICACSIS47736.2019.8979741

Baldwin, J., & Dong, X. (2004). *Simple Fuzzy Logic Rules Based on Fuzzy Decision Tree for Classification and And Prediction Problem*. *November 2014*.

Baldwin, J., & Karale, S. B. (2003). *New Concepts for Fuzzy Partitioning, Defuzzification and Derivation of Probabilistic Fuzzy Decision Trees*. 484–487.

Bataineh, A. Al. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. *International Journal of Machine Learning and Computing*, *9*(3), 248–254. https://doi.org/10.18178/ijmlc.2019.9.3.794

Begenova, S. B., & Avdeenko, T. V. (2018a). Building of fuzzy decision trees using ID3 algorithm. *Journal of Physics: Conference Series*, *1015*(2). https://doi.org/10.1088/1742-6596/1015/2/022002

Begenova, S. B., & Avdeenko, T. V. (2018b). The research of fuzzy decision trees building based on entropy and the theory of fuzzy sets. *CEUR Workshop Proceedings*, *2212*, 296–303. https://doi.org/10.18287/1613-0073-2018-2212-296-303

Benyoucef, L., & Grabot, B. (2013). *Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management.* Springer London.

Bertelsen, R., & Martinez, T. R. (1994). Extending ID3 Through Discretization of Continuous Inputs. *In Proceedings of the 7th Florida Artificial Intelligence Research Symposium*, 122–125.

Bethapudi, P., Reddy, E. S., & Varma, K. V. (2015). Classification of Breast Cancer using Gini Index based Fuzzy Supervised Learning in Quest Decision Tree Algorithm. *International Journal of Computer Applications*, *111*(14), 50–57. https://doi.org/10.5120/19610-1494

Bhadgale, P. A. M., & Natu, M. S. (2017). Implementation of Improved ID3 Algorithm Based on Association Function. *International Journal of Pure and Applied Mathematics*, *114*(10), 1–9.

Bilbao, I., & Bilbao, J. (2017). Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 173–177.

Breiman, L. (1996a). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1023/A:1018054314350

Breiman, L. (1996b). Bias, Variance, and Arcing Classifiers. In *Statistics Department, University of California at Berkeley, Berkeley, CA* (Vol. 53, Issue 1).

Brito, P. M. De. (2018). *Predicting the occurrence of breast cancer using insulin-related biomarkers , independently of obesity*. Tilburg University.

Campbell, M. J., & Swinscow, T. D. V. (2009). Statistics at Square One, 11th Edition. In *Wiley-Blackwell*. John Wiley & Sons. https://doi.org/10.1136/bmj.2.6042.1008-b

Chaurasia, V., & Pal, S. (2020). Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer. *SN Computer Science*, *1*(5). https://doi.org/10.1007/s42979-020-00296-8

Chen, Jie, Wang, X., & Zhai, J. (2009). Pruning Decision Tree Using Genetic Algorithms. *2009 International Conference on Artificial Intelligence and Computational Intelligence*, *3*, 244–248. https://doi.org/10.1109/AICI.2009.351

Chen, Jin, Luo, D., & Mu, F. (2009). An Improved ID3 Decision Tree Algorithm. *Proceedings of 2009 4th International Conference on Computer Science & Education*, 127–130. https://doi.org/10.1109/iccse.2009.5228509

Chen, M. S., & Wang, S. W. (1999). Fuzzy clustering analysis for optimizing fuzzy membership functions. *Fuzzy Sets and Systems*, *103*(2), 239–254. https://doi.org/10.1016/S0165-0114(98)00224-3

Chen, Y., Wang, T., Wang, B., & Li, Z. (2019). *A Survey of Fuzzy Decision Tree Classifier*. *8658*(2009). https://doi.org/10.1007/s12543-009-0012-2

Chiang, Y. M., Chang, L. C., & Chang, F. J. (2004). Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *Journal of Hydrology*, *290*(3–4), 297–311. https://doi.org/10.1016/j.jhydrol.2003.12.033

Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F., & Kløve, B. (2018). River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Science of the Total Environment*, *615*, 272–281. https://doi.org/10.1016/j.scitotenv.2017.09.293

Cintra, Marcos E., Camargo, H. A., & Martin, T. (2009). Optimising the fuzzy granulation of attribute domains. *2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference, IFSA-EUSFLAT 2009 - Proceedings*, *May 2014*, 742–747.

Cintra, Marcos E., Monard, M. C., Cherman, E. A., & De Arruda Camargo, H. (2011). On the estimation of the number of fuzzy sets for fuzzy rule-based classification systems. *Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems, HIS 2011*, *December*, 211–216. https://doi.org/10.1109/HIS.2011.6122107

Cintra, Marcos E, Meira, C. A. A., Monard, M. C., Camargo, H. A., & Rodrigues, L. H. A. (2011). The use of fuzzy decision trees for coffee rust warning in Brazilian crops. *2011 11th International Conference on Intelligent Systems Design and Applications*, 1347–1352.

Cintra, Marcos E, Monard, M. C., & Camargo, H. A. (2013). A Fuzzy Decision Tree Algorithm Based on C4.5. *Mathware & Soft Computing*, *20*, 56–62.

Cintra, Marcos E, & Neves, O. A. (2013). A Fuzzy Decision Tree for Bus Network Management. *Knowledge Discovery, Mining and Learning – KDMiLe*, 1–8.

Cintra, Marcos Evandro. (2012). *Genetic generation of fuzzy knowledge bases: new perspectives*. https://doi.org/10.11606/T.55.2012.TDE-16072012-144620

Cintra, Marcos Evandro, Francisco, J., Ribeiro, F., & Neves, O. A. (2014). A Fuzzy Decision Tree Model to Support the Task of Bus Reallocation in Public Transport Systems. *Journal of Information and Data Management*, *5*(1), 104–113.

Cintra, Marcos Evandro, Martin, T. P., Monard, M. C., & Camargo, H. de A. (2011). An Approach for the Extraction of Classification Rules from Fuzzy Formal Contexts. In *RELATÓRIOS TÉCNICOS DO ICMC TECHNICAL REPORTS*. https://web.icmc.usp.br/SCATUSU/RT/BIBLIOTECA_113_RT_368.pdf

Dai, X., Gao, L., & Dong, C. (2010). SELF-ADAPTIVE FUZZIFICATION IN FUZZY DECISION TREE INDUCTION. *2010 International Conference on Machine Learning and Cybernetics*, *1*, 296–301. https://doi.org/10.1109/ICMLC.2010.5581048

De Reus, N. M. (1994). Assessment of benefits and drawbacks of using fuzzy logic, especially in fire control systems. In *TNO Defense Research*. https://doi.org/10.1088/0031-9112/38/3/008

Desuky, A. S. (2018). Severity of breast masses prediction in mammograms based on optimized naive bayes diagnostic system. *Advances in Systems Science and Applications*, *18*(1), 12–19. https://doi.org/10.25728/assa.2018.18.1.266

Domingo, M. J., Gerardo, B. D., & Medina, R. P. (2019). Fuzzy decision tree for breast cancer prediction. *AISS '19: Proceedings of the International Conference on Advanced Information Science and System*, 1–6. https://doi.org/10.1145/3373477.3373489

Domingos, P. (2000). Bayesian Averaging of Classifiers and the Overfitting Problem. In *Icml*.

Douangnoulack, P., & Boonjing, V. (2018). Building Minimal Classification Rules for Breast Cancer Diagnosis. *2018 10th International Conference on Knowledge and Smart Technology (KST)*, 278–281.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, *35*(5–6), 352–359. https://doi.org/10.1016/S1532-0464(03)00034-0

Durgalakshmi, B., & Vijayakumar, V. (2020). Feature selection and classification using support vector machine and decision tree. *Computational Intelligence*, *August*, 1–13. https://doi.org/10.1111/coin.12280

Einipour, A. (2011). A Fuzzy-ACO Method for Detect Breast Cancer. *Global Journal of Health Science*, *3*(2), 195–199. https://doi.org/10.5539/gjhs.v3n2p195

El_Rahman, S. A. (2020). Predicting breast cancer survivability based on machine learning and features selection algorithms: a comparative study. In *Journal of Ambient Intelligence and Humanized Computing* (Issue January 2016). Springer Berlin Heidelberg. https://doi.org/10.1007/s12652-020-02590-y

Elaidi, H., Benabbou, Z., & Abbat, H. (2018). A comparative study of algorithms constructing decision trees : ID3. *Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications - LOPAL '18*, 3–7.

Elyan, E., & Gaber, M. M. (2016). A fine-grained Random Forests using class decomposition: an application to medical diagnosis. *Neural Computing and Applications*, *27*(8), 2279–2288. https://doi.org/10.1007/s00521-015-2064-z

Fajfer, M., & Janikow, C. Z. (2000). Bottom-up fuzzy partitioning in fuzzy decision trees. *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 326–330. https://doi.org/10.1109/nafips.2000.877446

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., & Weber, M. H. (2017). Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environmental Monitoring and Assessment*, *189*(7). https://doi.org/10.1007/s10661-017-6025-0

Gentle, J. E., Härdle, W. K., & Mori, Y. (2012). Handbook of computational statistics: Concepts and methods: Second Edition. In *Handbook of Computational Statistics: Concepts and Methods: Second Edition* (Issue January 2012). https://doi.org/10.1007/978-3-642-21551-3

Ghosh, A., Manwani, N., & Sastry, P. S. (2017). On the robustness of decision tree learning under label noise. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10234 LNAI*(May 2016), 685–697. https://doi.org/10.1007/978-3-319-57454-7_53

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Harwati, & Sudiya, A. (2016). Application of Decision Tree Approach to Student Selection Model- A Case Study. *IOP Conference Series: Materials Science and Engineering*, *105*(1). https://doi.org/10.1088/1757-899X/105/1/012014

Hassan, F., Shaheen, M. E., & Sahal, R. (2020). Real-time healthcare monitoring system using online machine learning and spark streaming. *International Journal*

of Advanced Computer Science and Applications, *11*(9), 650–658. https://doi.org/10.14569/IJACSA.2020.0110977

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf

Hazra, S., & Sanyal, S. (2016). Recruitment Prediction Using Id3 Decision Tree. *International Journal of Advance Engineering and Research Development*, *3*(10), 48–52. https://doi.org/10.21090/ijaerd.031010

Higa, A. (2018). Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms. *International Journal of Computer Applications Technology and Research*, *7*(1), 23–27. https://doi.org/10.7753/ijcatr0701.1004

Hssina, B., Merbouha, A., Ezzikouri, H., & Erritali, M. (2014). A comparative study of decision tree ID3 and C4 . 5. *International Journal of Advanced Computer Science and Applications*, *Special Issue on Advances in Vehicular Ad Hoc Networking and Applications(2)*. https://doi.org/10.14569/SpecialIssue.2014.040203

Hsu, K. W., & Srivastava, J. (2012). Improving bagging performance through multi-algorithm ensembles. *Frontiers of Computer Science in China*, *6*(5), 498–512. https://doi.org/10.1007/s11704-012-1163-6

Islam, M. M., Yao, X., Nirjon, S. M. S., Islam, M. A., & Murase, K. (2008). Bagging and boosting negatively correlated neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *38*(3), 771–784. https://doi.org/10.1109/TSMCB.2008.922055

Jacob, S. G., & Geetha Ramani, R. (2012). Efficient classifier for classification of prognostic breast cancer data through data mining techniques. *Lecture Notes in Engineering and Computer Science*, *1*, 493–498.

Jain, A. K., & Mao, J. (1996). Artificial Neural Network: A Tutorial. *Communications*, *29*, 31–44. https://doi.org/10.1109/2.485891

Janghel, R. R., Shukla, A., Tiwari, R., & Kala, R. (2010). Breast Cancer diagnosis using Artificial Neural Network models. *Proceedings - 3rd International Conference on Information Sciences and Interaction Sciences, ICIS 2010*, *Cl*, 89–94. https://doi.org/10.1109/ICICIS.2010.5534716

Janikow, C. Z. (1998). Fuzzy Decision Trees : Issues and Methods. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B (CYBERNETICS)*, *28*(1), 1–14.

Janitza, S., & Hornung, R. (2018). On the overestimation of random forest ’ s out-of-bag error. *PLoS ONE*, *13*(8), 1–31. https://doi.org/10.1371/journal.pone.0201904 August

Jearanaitanakij, K. (2005). Classifying continuous data set by ID3 algorithm. *2005 Fifth International Conference on Information, Communications and Signal Processing*, *2005*, 1048–1051. https://doi.org/10.1109/icics.2005.1689212

Jensen, W. A. (2008). Decision Trees for Business Intelligence and Data Mining: Using SAS® Enterprise Miner™. *Technometrics*, *50*(3), 409–410. https://doi.org/10.1198/tech.2008.s904

Jeyasingh, S., & Veluchamy, M. (2017). Modified bat algorithm for feature selection with the Wisconsin Diagnosis Breast Cancer (WDBC) dataset. *Asian Pacific Journal of Cancer Prevention*, *18*(5), 1257–1264. https://doi.org/10.22034/APJCP.2017.18.5.1257

Jiao, S. R., Song, J., & Liu, B. (2020). A Review of Decision Tree Classification Algorithms for Continuous Variables. *Journal of Physics: Conference Series*, *1651*(1), 1–6. https://doi.org/10.1088/1742-6596/1651/1/012083

Jin, C., Li, F., & Li, Y. (2014). A generalized fuzzy ID3 algorithm using generalized information entropy. *Knowledge-Based Systems*, *64*, 13–21. https://doi.org/10.1016/j.knosys.2014.03.014

Johra, F. T., & Shuvo, M. M. H. (2017). Detection of breast cancer from histopathology image and classifying benign and malignant state using fuzzy logic. *2016 3rd International Conference on Electrical Engineering and Information and Communication Technology, ICEEiCT 2016*, *1*, 2–6. https://doi.org/10.1109/CEEICT.2016.7873137

Jothikumar, R., & Siva Balan, R. (2016). C4 . 5 classification algorithm with back-track pruning for accurate prediction of heart disease . *Biomedical Research*, *Special Section: Computational Life Science and Smarter Technological Advancement*, 107–111.

Juneja, K., & Rana, C. (2020). An improved weighted decision tree approach for breast cancer prediction. *International Journal of Information Technology (Singapore)*, *12*(3), 797–804. https://doi.org/10.1007/s41870-018-0184-2

Kabiraj, S., Akter, L., Raihan, M., DIba, N. J., Podder, E., & Hassan, M. M. (2020). Prediction of Recurrence and Non-recurrence Events of Breast Cancer using Bagging Algorithm. *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*. https://doi.org/10.1109/ICCCNT49239.2020.9225440

Kantarci-Savas, S., & Nasibov, E. (2017). Fuzzy ID3 algorithm on linguistic dataset by using WABL defuzzification method. *IEEE International Conference on Fuzzy Systems*. https://doi.org/10.1109/FUZZ-IEEE.2017.8015502

Kavitha, K. K., Kangaiammal, A., & Satheesh, K. (2015). *Analysis On Classification Techniques In Mammographic Mass Data Set*. *5*(7), 32–35.

Kayaalp, F., & Basarslan, M. S. (2019). Performance Analysis Of Filter Based Feature Selection Methods On Diagnosis Of Breast Cancer And Orthopedics. *6th International Congress on Fundamental and Applied Sciences 2019 (ICFAS2019)*, *July*, 1–11.

Kerdprasop, N., & Kerdprasop, K. (2011). Discrete decision tree induction to avoid overfitting on categorical data. *Recent Researches in Computational Techniques,*

Non-Linear Systems and Control - Proc. of the 13th WSEAS Int. Conf. on MAMECTIS'11, NOLASC'11, CONTROL'11, WAMUS'11, *July 2011*, 247–252.

Khan, M. U., Choi, J. P., Shin, H., & Kim, M. (2008). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology," October 2015*, 5148–5151. https://doi.org/10.1109/iembs.2008.4650373

Khan, R. A., Suleman, T., Farooq, M. S., Rafiq, M. H., & Tariq, M. A. (2017). Data mining algorithms for classification of diagnostic cancer using genetic optimization algorithms. *Ijcsns*, *17*(12), 207.

Kharya, S., & Soni, S. (2016). Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection. *International Journal of Computer Applications*, *133*(9), 32–37. https://doi.org/10.5120/ijca2016908023

Khoshgoftaar, T. M., Hulse, J. Van, & Napolitano, A. (2011). Comparing Boosting and Bagging Techniques With Noisy and Imbalanced Data. *EEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS*, *41*(3), 552–568.

Kocsis, L., György, A., & Bán, A. N. (2013). BoostingTree: Parallel selection of weak learners in boosting, with application to ranking. *Machine Learning*, *93*(2–3), 293–320. https://doi.org/10.1007/s10994-013-5364-5

Kowsari, K., Bari, N., Vichr, R., & Goodarzi, F. A. (2018). *FSL-BM : Fuzzy Supervised Learning with Binary Meta-Feature for Classification*. *April*.

Kukreja, H., Bharath, N., Siddesh, C. S., & Kuldeep, S. (2016). AN INTRODUCTION TO ARTIFICIAL NEURAL NETWORK. *International Journal Of Advance Research And Innovative Ideas In Education*, *1*(5).

Kumar, M., & Rath, S. K. (2014). *Classification of Microarray Data Using Kernel Fuzzy Inference System*. *2014*. https://doi.org/10.1155/2014/769159

Kumar, S. S., & Nadu, T. (2013). Modified C4.5 Algorithm with Improved Information Entropy and Gain Ratio. *International Jurnal of Engineering Research & Technology (IJERT)*, *2*(9), 2768–2773.

Kumari, M., Singh, V., & Ahlawat, P. (2019). Decision Support System for Breast Cancer Prediction. *International Journal for Research in Applied Science and Engineering Technology*, *7*(3), 816–821. https://doi.org/10.22214/ijraset.2019.3142

Kuo, S. J., Hsiao, Y. H., Huang, Y. L., & Chen, D. R. (2008). Classification of benign and malignant breast tumors using neural networks and three-dimensional power Doppler ultrasound. *Ultrasound in Obstetrics and Gynecology*, *32*(1), 97–102. https://doi.org/10.1002/uog.4103

Kusuma, E. J., Shidik, G. F., & Pramunendar, R. A. (2020). Optimization of Neural Network using Nelder Mead in Breast Cancer Classification. *International Journal of Intelligent Engineering and Systems*, *13*(6), 330–337.

https://doi.org/10.22266/ijies2020.1231.29

Kwon, H., Park, J., & Lee, Y. (2019). Stacking ensemble technique for classifying breast cancer. *Healthcare Informatics Research*, *25*(4), 283–288. https://doi.org/10.4258/hir.2019.25.4.283

Lavanya, D., & Rani, K. U. (2011). Performance Evaluation of Decision Tree Classifiers on Medical Datasets. *International Journal of Computer Applications*, *26*(4), 1–4. https://doi.org/10.5120/3095-4247

Lavanya, D., & Usha Rani, K. (2012a). Ensemble Decision Tree Classifier For Breast Cancer Data. *International Journal of Information Technology Convergence and Services*, *2*(1), 17–24. https://doi.org/10.5121/ijitcs.2012.2103

Lavanya, D., & Usha Rani, K. (2012b). ENSEMBLE DECISION TREE CLASSIFIER FOR BREAST CANCER DATA. *International Journal of Information Technology Convergence and Services (IJITCS)*, *2*(1), 12–24. https://doi.org/10.5121/ijitcs.2012.2103

Lee, H. B., & Han, W. (2014). Unique features of young age breast cancer and its management. *Journal of Breast Cancer*, *17*(4), 301–307. https://doi.org/10.4048/jbc.2014.17.4.301

Li, Y., Jiang, D., & Li, F. (2012). The application of generating fuzzy ID3 algorithm in performance evaluation. *Procedia Engineering*, *29*, 229–234. https://doi.org/10.1016/j.proeng.2011.12.699

Liao, T. W., Celmins, A. K., & Hammell, R. J. (2003). A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems*, *135*(2), 241–257. https://doi.org/10.1016/S0165-0114(02)00136-7

Liew, X. Y., Hameed, N., & Clos, J. (2021). An investigation of XGBoost-based algorithm for breast cancer classification. *Machine Learning with Applications*, *6*(August), 100154. https://doi.org/10.1016/j.mlwa.2021.100154

Ma, Z., Wang, P., Gao, Z., Wang, R., & Khalighi, K. (2018). Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLoS ONE*, *13*(10), 1–12. https://doi.org/10.1371/journal.pone.0205872

Majeed, W., Aslam, B., Javed, I., Khaliq, T., Muhammad, F., Ali, A., & Raza, A. (2014). Breast Cancer: Major Risk Factors and Recent Developments in Treatment. *Asian Pacific Journal of Cancer Prevention*, *15*, 3353–3358. https://doi.org/https://doi.org/10.7314/APJCP.2014.15.8.3353

Marco, R., Herman, N. S., & Ahmad, S. S. S. (2019). Optimizing Software Effort Estimation Models Based On Metaheuristic Methods: A Proposed Framework. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(1.5), 294–304.

Markechová, D., & Rieˇcan, B. (2016). Entropy of Fuzzy Partitions and Entropy of Fuzzy Dynamical Systems. *Entropy*, *18*(1), 1–10.

https://doi.org/10.3390/e18010019

Matteo, R., & Giorgio, V. (2001). Ensemble methods: a review. In *Data Mining and Machine Learning for Astronomical Applications*.

Minivishapavithrathi, V., & Rani, K. S. (2019). Prediction of breast cancer using machine learning. *International Journal of Advanced Science and Technology*, *28*(7), 150–153. https://doi.org/10.2174/2213275912666190617160834

Modi, N., & Ghanchi, K. (2016). A comparative analysis of feature selection methods and associated machine learning algorithms on Wisconsin Breast Cancer dataset (WBCD). *Advances in Intelligent Systems and Computing*, *408*, 215–224. https://doi.org/10.1007/978-981-10-0129-1_23

Mohammed, M., Mwambi, H., Omolo, B., & Elbashir, M. K. (2018). Using stacking ensemble for microarray-based cancer classification. *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering, ICCCEEE 2018*, *December*, 1–8. https://doi.org/10.1109/ICCCEEE.2018.8515872

Mostafa, M. G., Perkins, T. C., & Farag, A. A. (2000). *SUPERVISED FUZZY AND BAYESIAN CLASSIFICATION O F HIGH DIMENSIONAL DATA : A COMPARATIVE STUDY*. 772–775.

Munther, A., Othman, R. R., Alalousi, A., Anbar, M., & Nizam, S. (2014). Network Traffic Classification - A Comparative Study of Two Common Decision Tree Methods : C4 . 5 and Random Forest. *2nd International Conference on Electronic Design (ICED)*.

Murty, M. R., Raju, S. K., Rao, M. V., & Satapaty, S. C. (2018). Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction. *International Journal of Pure and Applied Mathematics*, *118*(July), 321–334.

Mushtaq, Z., Yaqub, A., Sani, S., & Khalid, A. (2020). Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets. *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A*, *43*(1), 80–92. https://doi.org/10.1080/02533839.2019.1676658

Nai-Arun, N., & Punnee, S. (2014). Ensemble learning model for diabetes classification. *Advanced Materials Research*, *931–932*(May 2014), 1427–1431. https://doi.org/10.4028/www.scientific.net/AMR.931-932.1427

Naji, M. A., Filali, S. El, Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., & Debauche, O. (2021). Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis. *Procedia Computer Science*, *191*, 487–492. https://doi.org/10.1016/j.procs.2021.07.062

Nguyen, T., Hettiarachchi, I., Khosravi, A., Salaken, S. M., & Bhatti, A. (2017). *Multiclass EEG Data Classification using Fuzzy Systems*. 0–5.

Nijhawan, V. K., Madan, M., & Dave, M. (2017). The Analytical Comparison of ID3 and C4 . 5 using WEKA. *International Journal of Computer Applications*, *167*(11), 1–4.

Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, *34*(4), 133–144. https://doi.org/10.1016/j.tele.2017.01.007

Nseerullah. (2018). THE PROS AND CONS OF PRUNING IN CLASSIFICATION. *Proceedings of Academicsera 32nd International Conference, London, United Kingdom*, 24–25. https://doi.org/10.1093/asj/sjx265

Odegua, R. O. (2020). An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking). *Deep Learning IndabaX*, *12*(10), 1683.

Olaru, C., & Wehenkel, L. (2003). A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, *138*(2), 221–254. https://doi.org/10.1016/S0165-0114(03)00089-7

Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine Learning Classification Techniques for Breast Cancer Diagnosis. *IOP Conference Series: Materials Science and Engineering*, *495*(1). https://doi.org/10.1088/1757-899X/495/1/012033

Ortiz-Rodriguez, J. M., Guerrero-Mendez, C., Martinez-Blanco, M. del R., Castro-Tapia, S., Moreno-Lucio, M., Jaramillo-Martinez, R., Solis-Sanchez, L. O., Martinez-Fierro, M. de la L., Garza-Veloz, I., Galvan, J. C. M., & Garcia, J. A. B. (2018). Breast Cancer Detection by Means of Artificial Neural Networks. In *Advanced Applications for Artificial Neural Networks*. https://doi.org/10.5772/intechopen.71256

Osisanwo, F. ., Akinsola, J. E. ., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138. https://doi.org/10.14445/22312803/ijctt-v48p126

Oyelade, O. N., Obiniyi, A. A., Junaidu, S. B., & Adewuyi, S. A. (2018). ST-ONCODIAG: A semantic rule-base approach to diagnosing breast cancer base on Wisconsin datasets. *Informatics in Medicine Unlocked*, *10*(December 2017), 117–125. https://doi.org/10.1016/j.imu.2017.12.008

Padhi, T., & Kumar, P. (2019). Breast cancer analysis using WEKA. *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, 229–233. https://doi.org/10.1109/CONFLUENCE.2019.8776911

Patel, B. N., Prajapati, S. G., & Lakhtaria, K. I. (2012). Efficient Classification of Data Using Decision Tree. *Bonfring International Journal of Data Mining*, *2*(1), 06–12. https://doi.org/10.9756/bijdm.1098

Patil., M. S., Agrawal, M. . M., & Baviskar, M. V. R. (2015). Efficient Processing of Decision Tree Using ID3 & improved C4.5 Algorithm. *International Journal of Computer Science and Information Technologies*, *6*(2), 1956–1961.

Patrício, M., Pereira, J., Crisóstomo, J., Matafome, P., Gomes, M., Seiça, R., & Caramelo, F. (2018). Using Resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*, *18*(1), 1–8. https://doi.org/10.1186/s12885-017-

3877-1

Peña-Reyes, C. A., & Sipper, M. (2001). Fuzzy CoCo: A cooperative-coevolutionary approach to fuzzy modeling. *IEEE Transactions on Fuzzy Systems*, *9*(5), 727–737. https://doi.org/10.1109/91.963759

Petersen, M. L., Molinaro, A. M., Sinisi, S. E., & van der Laan, M. J. (2007). Cross-validated bagged learning. *Journal of Multivariate Analysis*, *98*(9), 1693–1704. https://doi.org/10.1016/j.jmva.2007.07.004

Petri, C. (2010). *Decision Trees*. https://doi.org/10.1201/9781420090741.ch6

Piramuthu, S. (2008). Input data for decision trees. *Expert Systems with Applications*, *34*(2), 1220–1226. https://doi.org/10.1016/j.eswa.2006.12.030

Prasetyo, C., Kardiana, A., & Yuliwulandari, R. (2014). Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques. *International Journal of Advanced Research in Artificial Intelligence*, *3*(7), 10–14. https://doi.org/10.14569/ijarai.2014.030703

Prati, R. C., Charte, F., & Herrera, F. (2017). *A first approach towards a fuzzy decision tree for multilabel classification*. *July*. https://doi.org/10.1109/FUZZ-IEEE.2017.8015521

Pulkkinen, P., & Koivisto, H. (2010). A dynamically constrained multiobjective genetic fuzzy system for regression problems. *IEEE Transactions on Fuzzy Systems*, *18*(1), 161–177. https://doi.org/10.1109/TFUZZ.2009.2038712

Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., & Subbotin, S. (2019). Application of Fuzzy Decision Tree for Signal Classification. *IEEE Transactions on Industrial Informatics*, *15*(10), 5425–5434. https://doi.org/10.1109/TII.2019.2904845

Rajesh, K. and S. A. (2012). Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, *1*(2), 72–77. www.ijarcce.com

Ramanan, A., Suppharangsan, S., & Niranjan, M. (2007). Unbalanced decision trees for multi-class classification. *ICIIS 2007 - 2nd International Conference on Industrial and Information Systems 2007, Conference Proceedings*, *September*, 291–294. https://doi.org/10.1109/ICIINFS.2007.4579190

Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, *29*(1), 65–74. https://doi.org/10.1016/j.eswa.2005.01.006

Rishika, V., & Sowjanya, A. M. (2019). Prediction of breast cancer using ensemble learning. *International Journal of Management, Technology And Engineering*, *IX*(1), 1857–1867. https://doi.org/10.1109/ICAEE48663.2019.8975544

Sahri, Z., Arif, F., Sakinah, S., Ahmad, S., Ahmad, R., & Yusof, R. (2017). *Imputing*

Missing Values in Mammography Mass Dataset: Will it Increase Classification Performance of Machine Learning Algorithms? 8–12. https://doi.org/10.17758/eirai.iae1017305

Sakib, S., Siddique, M. A. B., & Rahman, M. A. (2020). Performance Evaluation of t-SNE and MDS Dimensionality Reduction Techniques with KNN , ENN and SVM Classifiers. *2020 IEEE Region 10 Symposium (TENSYMP)*, *June*, 5–7.

Salehi-Moghaddami, N., Yazdi, H. S., & Poostchi, H. (2011). Correlation based splitting criterionin multi branch decision tree. *Central European Journal of Computer Science*, *1*(2), 205–220. https://doi.org/10.2478/s13537-011-0017-x

Santoso, H. B. (2020). Fuzzy Decision Tree to Predict Student Success in Their Studies. *International Journal of Quantitative Research and Modeling*, *1*(3), 135–144. https://doi.org/10.46336/ijqrm.v1i3.59

Saoud, H., Ghadi, A., Ghailani, M., & Abdelhakim, B. A. (2019). *Using Feature Selection Techniques to Improve the Accuracy of Breast Cancer Classification* (Issue January). Springer International Publishing. https://doi.org/10.1007/978-3-030-11196-0_28

Sardouk, F., Duru, A. D., & Bayat, O. (2019). Classification of breast cancer using data mining. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, *51*(1), 38–46.

Sathiyanarayanan, P., Pavithra, S., Sai Saranya, M., & Makeswari, M. (2019). Identification of breast cancer using the decision tree algorithm. *2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019*, 1–6. https://doi.org/10.1109/ICSCAN.2019.8878757

Schaefer, G., Nakashima, T., Závišek, M., Yokota, Y., Drastich, A., & Ishibuchi, H. (2007). Breast cancer classification using statistical features and fuzzy classification of thermograms. *IEEE International Conference on Fuzzy Systems*. https://doi.org/10.1109/FUZZY.2007.4295520

Schapire, R. E. (1990). The Strength of Weak Learnability. *Machine Learning*, *5*(2), 197–227. https://doi.org/10.1023/A:1022648800760

Schapire, R. E., & Freund, Y. (1995). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*, 119–139. https://doi.org/10.1145/2818346.2823306

Sehgal, G., Rawat, M., Gupta, B., Gupta, G., Sharma, G., & Shroff, G. (2018). Visual Predictive Analytics using iFuseML. *EuroVisWorkshop on Visual Analytics*. https://doi.org/10.2312/eurova.20181106

Senkamalavalli, R., & Bhuvaneswari, T. (2018). Improved Classification of Breast Cancer Data Using Hybrid Techniques. *IInternational Journal of Advanced Engineering Research and Science (IJAERS)*, *5*(5), 77–81. https://doi.org/10.22161/ijaers.5.5.11

Seymoens, T., Ongenae, F., Jacobs, A., Verstichel, S., & Ackaert, A. (2019). A

methodology to involve domain experts and machine learning techniques in the design of human-centered algorithms. In *IFIP Advances in Information and Communication Technology* (Vol. 544). Springer International Publishing. https://doi.org/10.1007/978-3-030-05297-3_14

Shah, R., Rosso, K., & David Nathanson, S. (2014). Pathogenesis, prevention, diagnosis and treatment of breast cancer. *World Journal of Clinical Oncology*, *5*(3), 283–298. https://doi.org/10.5306/wjco.v5.i3.283

Sharifmoghadam, M., & Jazayeriy, H. (2019). Breast Cancer Classification Using AdaBoost- Extreme Learning Machine. *5th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2019*, *December*, 1–5. https://doi.org/10.1109/ICSPIS48872.2019.9066088

Sharma, H., & Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal of Science and Research (IJSR)*, *5*(4), 2094–2097. https://doi.org/10.21275/v5i4.nov162954

Shimojima, K., Fukuda, T., & Hasegawa, Y. (1995). Self-tuning fuzzy modeling with adaptive membership function, rules, and hierarchical structure based on genetic algorithm. *Fuzzy Sets and Systems*, *71*(3), 295–309. https://doi.org/10.1016/0165-0114(94)00280-K

Singh, S., & Manoj, G. (2014). Comparative Study Id3, Cart and C4.5 Decision Tree Algorithm: a Survey. *International Journal of Advanced Information Science and Technology*, *27*(27), 97–103. https://doi.org/10.15693/ijaist/2014.v3i7.47-52

Sizilio, G. R. M. A., Leite, C. R. M., Guerreiro, A. M. G., & Neto, A. D. D. (2012). Fuzzy method for pre-diagnosis of breast cancer from the Fine Needle Aspirate analysis. *BioMedical Engineering Online*, *11*, 1–21. https://doi.org/10.1186/1475-925X-11-83

Song, Q., Shepperd, M., Chen, X., & Liu, J. (2008). Can k-NN imputation improve the performance of C4.5 with small software project data sets? A comparative evaluation. *Journal of Systems and Software*, *81*(12), 2361–2370. https://doi.org/10.1016/j.jss.2008.05.008

Sree Kumar, S., Radhakrishnan, A. ., & Cheong, S. . (2010). Rapid Metastasis of Breast Cancer Cells from Primary Tumour to Liver. *Pakistan Journal of Biological Sciences*, *13*(7), 303–315.

Suca, C., Córdova, A., Condori, A., Jordy, C., & Sulla, J. (2016). Fuzzy Model for Prediction Childhood Obesity Cases Using the Generalized GFID3. *XVII Conferencia de La Asociación Española Para La Inteligencia Artificial*, *July*.

Suganya, X. J., & Balasubramanian, R. (2016). Enhanced ID3 algorithm based on the weightage of the Attribute. *International Journal of Innovatice Research in Advanced Engineering*, *3*(03), 12–18. http://www.ijirae.com/volumes/Vol3/iss3/03.MRAE10083.pdf

Sumbaly, R., Vishnusri, N., & Jeyalatha, S. (2014). Diagnosis of Breast Cancer using Decision Tree Data Mining Technique. *International Journal of Computer*

Applications, *98*(10), 16–24. https://doi.org/10.5120/17219-7456

Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., Shi, W., Jiang, J., Yao, P. P., & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. *International Journal of Biological Sciences*, *13*(11), 1387–1397. https://doi.org/10.7150/ijbs.21635

Surya Prasad, P., Tarun Varma, V., Sai, V., Harish, V., & Kumar, K. S. R. (2012). Classification of Different Textures Using SVM and Fuzzy logic. *International Journal of Advanced Computer Research*, *4*, 4–7.

Tahmasebi, P., & Hezarkhani, A. (2012). A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers and Geosciences*, *42*(April 2016), 18–27. https://doi.org/10.1016/j.cageo.2012.02.004

Thaker, S., & Nagori, V. (2018). Analysis of Fuzzification Process in Fuzzy Expert System. *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*, *132*, 1308–1316. https://doi.org/10.1016/j.procs.2018.05.047

Trees, F. D., & Fdt, A. (2004). *Fuzzy decision trees 2.1*. 16–23.

Tu, C., Liu, H., & Xu, B. (2017). AdaBoost typical Algorithm and its application research. *MATEC Web of Conferences*, *139*(00222). https://doi.org/10.1051/matecconf/201713900222

Tu, M. C., Shin, D., & Shin, D. (2009). A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms. *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, 183–187. https://doi.org/10.1109/DASC.2009.40

Umanol, M., Okamoto, H., Hatono, I., & Tamura, H. (1994). Fuzzy Decision Trees by Fuzzy ID3 Algorithm and Its Application to Diagnosis Systems. *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference*, *1*, 2113–2118. https://doi.org/10.1109/FUZZY.1994.343539

Venkatesan, E., & Velmurugan, T. (2015). Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. *Indian Journal of Science and Technology*, *8*(12), 83–89. https://doi.org/10.17485/ijst/2015/v8i29/84646

Verma, A. K., & Pal, S. (2020). Prediction of Skin Disease with Three Different Feature Selection Techniques Using Stacking Ensemble Method. *Applied Biochemistry and Biotechnology*, *191*(2), 637–656. https://doi.org/10.1007/s12010-019-03222-8

Vezhnevets, A., & Barinova, O. (2007). Avoiding boosting overfitting by removing confusing samples. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4701 LNAI*, 430–441. https://doi.org/10.1007/978-3-540-74958-5_40

Vrigazova, B. P. (2020). Detection of Malignant and Benign Breast Cancer Using the ANOVA-BOOTSTRAP-SVM. *Journal of Data and Information Science*, *5*(2), 62–75. https://doi.org/10.2478/jdis-2020-0012

Wang, L. X., & Mendel, J. M. (1992). Generating Fuzzy Rules by Learning from Examples. *IEEE Transactions on Systems, Man and Cybernetics*, *22*(6), 1414–1427. https://doi.org/10.1109/21.199466

Wang, X., & Wu, C. (2005). *USING IMPROVED SVM DECISION TREE TO CLASSIFY HRRP*. *August*, 18–21.

Wang, Y., Li, Y., Song, Y., Rong, X., & Zhang, S. (2017). Improvement of ID3 algorithm based on simplified information entropy and coordination degree. *Algorithms*, *10*(4), 1–18. https://doi.org/10.3390/a10040124

Watkins, E. J. (2019). Overview of breast cancer. *Journal of the American Academy of PAs*, *32*(10), 13–17. https://doi.org/10.1097/01.JAA.0000580524.95733.3d

Wu, J., & Hicks, C. (2021). Breast cancer type classification using machine learning. *Journal of Personalized Medicine*, *11*(2), 1–12. https://doi.org/10.3390/jpm11020061

Xiaosheng, L., & Bo, L. (2015). Recognition of Power Quality Disturbances Based on T-S Fuzzy Logic. *2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control*, *1*, 302–305. https://doi.org/10.1109/IMCCC.2015.70

Yang, S., Guo, J. Z., & Jin, J. W. (2018). An improved Id3 algorithm for medical data classification. *Computers and Electrical Engineering*, *65*, 474–487. https://doi.org/10.1016/j.compeleceng.2017.08.005

Yuan, Y., & Shaw, M. J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, *69*(2), 125–139. https://doi.org/10.1016/0165-0114(94)00229-Z

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions On*, *1*, 28–44. https://doi.org/10.1109/TSMC.1973.5408575

Zadeh, L. A. (2008). Is there a need for fuzzy logic? *Information Sciences*, *178*(13), 2751–2779. https://doi.org/10.1016/j.ins.2008.02.012

Zaitseva, E., & Levashenko, V. (2016). Construction of a reliability structure function based on uncertain data. *IEEE Transactions on Reliability*, *65*(4), 1710–1723. https://doi.org/10.1109/TR.2016.2578948

Zhai, J., Wang, X., Zhang, S., & Hou, S. (2018). Tolerance rough fuzzy decision tree. *Information Sciences*, *465*, 425–438. https://doi.org/10.1016/j.ins.2018.07.006

Zhao, L., Lee, S., & Jeong, S.-P. (2021). Decision Tree Application to Classification Problems with Boosting Algorithm. *Electronics*, *10*(1903). https://doi.org/https://doi.org/10.3390/electronics10161903

Zimmerman, R. K., Balasubramani, G. K., Nowalk, M. P., Eng, H., Urbanski, L., Jackson, M. L., Jackson, L. A., Mclean, H. Q., Belongia, E. A., Monto, A. S., Malosh, R. E., Gaglani, M., Clipper, L., Flannery, B., & Wisniewski, S. R. (2016). Classification and Regression Tree (CART) analysis to predict influenza in

primary care patients. *BMC Infectious Diseases*, 1–11. https://doi.org/10.1186/s12879-016-1839-x

Zupan, J. (1994). Introduction to Artificial Neural Network (ANN) Methods : What They Are and How to Use Them. *Acta Chimica Slovenica*, *41*(3), 327–352.