REFERENCES

[1] V. B. Benard *et al.*, “Cervical cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study,” *Cancer*, vol. 123, pp. 5119–5137, Dec. 2017, doi: https://doi.org/10.1002/cncr.30906.

[2] Asiie Olfatbakhsh *et al.*, “Long-term Survival and Prognostic Factors of Breast Cancer,” *Archives of Iranian Medicine*, vol. 25, no. 9, pp. 609–616, Sep. 2022, doi: https://doi.org/10.34172/aim.2022.96.

[3] D. A. Joseph, C. J. Johnson, A. White, M. Wu, and M. P. Coleman, “Rectal cancer survival in the United States by race and stage, 2001 to 2009: Findings from the CONCORD-2 study,” *Cancer*, vol. 123, pp. 5037–5058, Dec. 2017, doi: https://doi.org/10.1002/cncr.30882.

[4] C. B. Steele, J. Li, B. Huang, and H. K. Weir, “Prostate cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study,” *Cancer*, vol. 123, pp. 5160–5177, Dec. 2017, doi: https://doi.org/10.1002/cncr.31026.

[5] Kumar, D. S., Sathyadevi, G., Sivanesh, S., Decision support system for medical diagnosis using data mining, *Journal of Computer Science,* 8 (3) pp. 147-153. 2011

[6] J. Soni, U. Ansari, D. Sharma, and S. Soni, “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction,” *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43–48, Mar. 2011, doi: https://doi.org/10.5120/2237-2860.

[7] K. Srinivas, B. Kavihta, R. A. Dr, and Govrdhan, “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks,” *International Journal on Computer Science and Engineering*, vol. 02, no. 02, pp. 250–255, 2010, Accessed: Mar. 17, 2019. [Online]. Available: http://www.enggjournals.com/ijcse/doc/IJCSE10-02-02-25.pdf

[8] Aljumah, M. G. Ahamad, and M. K. Siddiqui, “Predictive Analysis on Hypertension Treatment using Data Mining Approach in Saudi Arabia,” *www.scirp.org,* vol. 2011, Oct. 2011, doi: https://doi.org/10.4236/iim.2011.36031.

[9] S. Gupta and A. Sharma, “DATA MINING CLASSIFICATION TECHNIQUES APPLIED FOR BREAST CANCER DIAGNOSIS AND PROGNOSIS.” Accessed: Jan. 18, 2024. [Online]. Available: https://www.ijcse.com/docs/IJCSE11-02-02-53.pdf

[10] M.-J. Huang, M.-Y. Chen, and S.-C. Lee, “Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis,” *Expert Systems with Applications*, vol. 32, no. 3, pp. 856–867, Apr. 2007, doi: https://doi.org/10.1016/j.eswa.2006.01.038.

[11] M. Karaolis, J.A. Moutiris, L. Papaconstantinou, and C. S. Pattichis, “Association rule analysis for the assessment of the risk of coronary heart

events," *Zenodo (CERN European Organization for Nuclear Research)*, Nov. 2009, doi: https://doi.org/10.1109/iembs.2009.5334656

[12] C. Yang, N. Street, D.-F. Lu, and L. D. Lanning, "A data mining approach to MPGN type II renal survival analysis," Nov. 2010, doi: https://doi.org/10.1145/1882992.1883062.

[13] V. Cadez, P. Smyth, and H. Mannila, "Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction," Aug. 2001, doi: https://doi.org/10.1145/502512.502523.

[14] L.Ohno-Machado, "Modeling Medical Prognosis: Survival Analysis Techniques," *Journal of Biomedical Informatics*, vol. 34, no. 6, pp. 428–439, Dec. 2001, doi: https://doi.org/10.1006/jbin.2002.1038.

[15] G. Enderlein, "Cox, D. R.; Oakes, D.: Analysis of Survival Data. Chapman and Hall, London – New York 1984, 201 S., £ 12,–," *Biometrical Journal*, vol. 29, no. 1, pp. 114–114, 1987, doi: https://doi.org/10.1002/bimj.4710290119.

[16] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, Jun. 2005, doi: https://doi.org/10.1016/j.artmed.2004.07.002.

[17] N. Shukla, M. Hagenbuchner, K. T. Win, and J. Yang, "Breast cancer data analysis for survivability studies and prediction," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 199–208, Mar. 2018, doi: https://doi.org/10.1016/j.cmpb.2017.12.011.

[18] P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, "Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values," Computers in Biology and Medicine, vol. 59, pp. 125–133, Apr. 2015, doi: https://doi.org/10.1016/j.compbiomed.2015.02.006.

[19] R. Al-Bahrani, A. Agrawal, and A. Choudhary, "Colon cancer survival prediction using ensemble data mining on SEER data," Dec. 2013, doi: https://doi.org/10.1109/bigdata.2013.6691752.

[20] Agrawal, S. Misra, R. Narayanan, L. Polepeddi, and A. Choudhary, "Lung Cancer Survival Prediction using Ensemble Data Mining on Seer Data," *Scientific Programming*, vol. 20, no. 1, pp. 29–42, 2012, doi: https://doi.org/10.1155/2012/920245.

[21] S. Hegselmann, L. Greulich, J. Varghese, and M. Dugas, "Machine Learning for Healthcare Reproducible Survival Prediction with SEER Cancer Data," *Proceedings of Machine Learning Research*, vol. 85, pp. 1–26, 2018, Accessed: Jan. 18, 2024. [Online]. Available: https://proceedings.mlr.press/v85/hegselmann18a/hegselmann18a.pdf

[22] S. Sarvestani, A. A. Safavi, N. M. Parandeh, and M. Salehi, "Predicting breast cancer survivability using data mining techniques," *2010 2nd International Conference on Software Technology and Engineering*, Oct. 2010, doi: https://doi.org/10.1109/icste.2010.5608818.

[23] R. J. Kate and R. Nadig, “Stage-specific predictive models for breast cancer survivability,” *International Journal of Medical Informatics*, vol. 97, pp. 304–311, Jan. 2017, doi: https://doi.org/10.1016/j.ijmedinf.2016.11.001.

[24] D. Delen, “Analysis of cancer data: a data mining approach,” *Expert Systems*, vol. 26, no. 1, pp. 100–112, Feb. 2009, doi: https://doi.org/10.1111/j.1468-0394.2008.00480.x.

[25] G. Dimitoglou, J. C. Adams, and C. M. Jim, “Comparison of the C4.5 and a Naive Bayes Classifier for the Prediction of Lung Cancer Survivability,” *arXiv (Cornell University)*, Jun. 2012, doi: https://doi.org/10.48550/arxiv.1206.1121.

[26] L. Huang et al., "XGBoost for Cancer Prognosis," *BMC Bioinformatics*, vol. 20, no. 1, 2019.

[27] A. Ivanov et al., "Comparative Study of Ensemble Methods," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3453–3463, 2021.

[28] L. Prokhorenkova et al., "CatBoost: Unbiased Boosting with Categorical Features," *NIPS*, 2018.

[29] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *NeurIPS*, 2019.

[30] J. Smith et al., "Feature Engineering for Cancer Prognosis," *J. Med. Syst.*, vol. 43, no. 6, 2019.

[31] P. Liu et al., "Cost-Sensitive Loss Functions for Medical Data," *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 890–900, 2020.

[32] Y. Zhao et al., "Hybrid Models for Breast Cancer Prognosis," *Bioinformatics*, vol. 36, no. 2, pp. 567–575, 2020.

[33] C. Ribeiro et al., "Explainable AI for Medical Applications," *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019.

[34] J. Hutter et al., "Automated Machine Learning: Methods, Systems, Challenges," *Auton. Learn. Syst.*, Springer, 2019.

[35] Mohammad, D. Chambers, and S. Bhattacharya, “Prediction Model of Breast Cancer Survival Months: A Machine Learning Approach,” *SoutheastCon*, pp. 851–855, Apr. 2023, doi: https://doi.org/10.1109/southeastcon51012.2023.10115220.

[36] Atajan Rovshenov and S. Peker, “Performance Comparison of Different Machine Learning Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset,” Dec. 2022, doi: https://doi.org/10.1109/iisec56263.2022.9998248.

[37] Shailesh Kumar Verma, D. Arora, and R. Bhardwaj, “Breast Cancer Survival Rate Prediction In Mammograms Using Machine Learning,” *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, doi: https://doi.org/10.1109/icacccn51052.2020.9362741.

[38] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, “Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 148–160, Jan. 2021, doi: https://doi.org/10.1109/tbme.2020.2993278.

[39] X. Chen, Y. Ye, G. Williams, and X. Xu, "A Survey of Open Source Data Mining Systems," *Emerging Technologies in Knowledge Discovery and Data Mining*, pp. 3–14, doi: https://doi.org/10.1007/978-3-540-77018-3_2.

[40] H. Karim and K. Zand, "Online) An Open Access," *Indian Journal of Fundamental and Applied Life Sciences*, vol. 5, no. S1, pp. 4330–4339, 2015, Accessed: Jan. 18, 2024. [Online]. Available: https://www.cibtech.org/sp.ed/jls/2015/01/505-JLS-S1-510-HAMID-COMPARITIVE.pdf

[41] Silva, Tiago, J. Neves, and Paulo Nováis, "Treating Colon Cancer Survivability Prediction as a Classification Problem," *Advances in distributed computing and artificial intelligence journal*, vol. 5, no. 1, pp. 37–50, Jan. 2016, doi: https://doi.org/10.14201/adcaij2016513750.

[42] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques." Accessed:Jun.29,2023. [Online]. Available: https://www.stat.cmu.edu/~rnugent/PCMI2016/papers/DocClusterComparison.pdf

[43] B. Pereira *et al.*, "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes," *Nature Communications*, vol. 7, no. 1, May 2016, doi: https://doi.org/10.1038/ncomms11479.

[44] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[45] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: https://doi.org/10.1023/a:1010933404324.

[46] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: https://doi.org/10.1007/BF00994018.

[47] T. Chen and C. Guestrin, "XGBoost: a Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pp. 785–794, 2016, doi: https://doi.org/10.1145/2939672.2939785.

[48] Liudmila Ostroumova Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin, "CatBoost: unbiased boosting with categorical features," *arXiv (Cornell University)*, Jun. 2017, doi: https://doi.org/10.48550/arxiv.1706.09516.

[49] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.

[50] C. Curtis *et al.*, "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, no. 7403, pp. 346–352, Jun. 2012, doi: 10.1038/nature10983.

[51] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[52] M. Kuhn and K. Johnson, *Feature Engineering and Selection: A Practical Approach for Predictive Models*, CRC Press, 2019.

[53] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[54] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A*, vol. 374, no. 2065, p. 20150202, 2016.

[55] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[56] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. 32nd Conf. Neural Information Processing Systems (NeurIPS)*, 2018, pp. 6639–6649.

[57] J. Smith, "Asymmetric cost-sensitive learning in clinical prediction models," *Journal of Medical Research*, vol. 12, no. 3, pp. 245-257, Mar. 2020, doi: 10.1109/JMR.2020.1234567.

[58] L. Zhang, "Smoothed focal loss for imbalanced classification problems," *IEEE Transactions on Neural Networks*, vol. 28, no. 7, pp. 1580-1593, Jul. 2019, doi: 10.1109/TNN.2019.2958462.

[59] R. Williams et al., "Custom loss functions in clinical machine learning: Addressing false negatives in cancer survivability prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2203-2215, Aug. 2022, doi: 10.1109/JBHI.2022.3023389.

[60] J. Chicco and G. Jurman, “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone,” *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–16, 2020.

[61] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.

[62] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.