Application of Machine Learning Algorithm for Breast Cancer Detection: A Thematic Analysis

Shahzma Tahseen[1*],Muhammad Ali Memon[2], Kamran Taj Pathan[2], Shah Muhammad Kamran[3], Waqar Ahmed Suthar[3]

***Abstract*- The application of AI and machine learning algorithms for timely breast cancer detection is gaining significant volume and researchers from medical science and computer science both are in search of a solution that can reduce the mortality rate through early detection and treatment. This study aims to assess different machine learning models with acceptable generalization capacity and clinical explainability based on two features; first to measure the association between independent variables (breast cancer risk and preventive factor) and dependent variable (existence of cancer). Secondly to explore the patterns and correlations in raw data, over time and case, and continuously learn and improve. For the collection of data, academic databases; ScienceDirect and PubMed were used to find the articles. These articles were then analyzed through NVivo for content analysis. The study conjectured that machine learning and artificial intelligence are helping the oncologist in the early detection of breast cancer which resultantly increases the survival rate of the patients. The above findings aim to stimulate the process of cancer scanning at an early stage and also point out the new horizon of health application to AI experts.**

***Keywords*— Breast Cancer, Machine Learning, Machine Learning based Algorithms, Thematic Analysis**

BACKGROUND

Cancer is one of the fastest-growing non-communicable illnesses, with 19.3 million new cases and 10 million deaths expected by 2020. Around the world, one out of every five men and one out of every six women is diagnosed with cancer, with one out of every eight men and one out of every eleven women dying from the disease. With 2.3 million new cases in 2020, female breast cancer is the most often diagnosed disease and a leading cause of cancer deaths, accounting for 11.7 percent of all new cancer cases. [1]. Breast cancer is a malignant cell growth that occurs abnormally in the breast. If not treated, it might spread to other bodyparts. Breast cancer, excluding skin cancer, is the most common cancer among women. [2].

Breast cancer diagnosis and treatment is an expensive and time-consuming process that costs a lot of money not only to the patients and their families but also to the government. Hence, developing BC preventive measures and tools for early detection is crucial [3]. Early detection of cancer needs less expensive and concise treatment, and it minimizes the risk of spreading to other body parts [4], [5]. Many international health organizations, such as the World Health Organization (WHO) and the International Alliance for Cancer Early Detection (IACED), promote early detection to reduce death rates [6].

Early breast cancer detection is even more difficult in developing countries [1] where the public and governments have fewer resources, and late cancer diagnosis leads to increased mortality. Many studies have accentuated the importance of screening in early cancer detection and survival; nevertheless, breast cancer screening programs have their unique characteristics and trade-offs, such as generic risk and screening approach, customized treatment strategy, and patient sampling. [7]–[9]. Furthermore, medical procedures such as diagnosis are difficult to make since they need the assessment of a variety of factors that may be competing, complimentary, or contradictory. These aspects interact with one another in both unilateral and bilateral ways, forming the foundation for overall therapeutic judgments [10].

These specificities and limits of cancer and screening need to be addressed with caution, and a generic approach is required for diagnosis, with little or no medical intervention. Statistical methods such as generalized linear models, regression, correlation, and others have been utilized and promoted for the detection and classification of these cancer [11]. Statistical approaches can also lay the groundwork for the use of artificial intelligence to improve diagnosis accuracy and precision.

Artificial Intelligence (AI) algorithms are assisting in the improvement of the screening process since they can analyze large amounts of multi-modal data and find signals that would otherwise be challenging to find. [12]–[14]. Artificial Intelligence (AI) educates machines to understand by

[1] LUMHS Jamshoro
[2] University of Sindh
[3] MUISTD-MUET Jamshoro
Country: Pakistan
Email: coordinator_it@lumhs.edu.pk

algorithm and make predictions based on past experiences. These learnings can aid cancer detection by defining criteria and automating the scientific procedure. Machine Learning (ML) is classified into two types: supervised learning, in which the system knows the outcome, and unsupervised learning, in which the system does not know the outcome. Both approaches use data to forecast the presence or absence of cancer, also the likelihood of cancer occurrence. The most often utilised algorithms in health sciences are Natural Language Processing (NLP), Convolutional Neural Network (CNN), and MLP. [15], [16].

For many years, ML algorithms have been employed to diagnose breast cancer. Algorithms use data patterns to forecast the likelihood of occurrence and classifies the types of cancer. In automated mode, ML employs a variety of probabilistic, statistical, and optimization modules to train and improve performance of detection system. ML algorithms may be applied to current and historical data, with expertise and without reprogramming or predefined instructions. After that, statistical methods and machine learning are used to examine the data. Machine learning algorithms can extract fundamental traits and potential rules that would be impossible to derive using statistics alone[2].

Machine and deep learning (D&ML)-based AI research and applications are now increasing. D&ML approaches are employed in the health sciences to analyze imaging and pathology reports, patient records, and other associated data [17], [18]. In addition, ML has been used in several studies to identify and predict breast cancer using data from risk and preventative variables [19]–[22].

This study aims to review different machine learning models with acceptable generalization ability and medical explainability based on two capabilities: first, to measure the association between independent variables (breast cancer risk and preventive factor) and dependent variable (breast cancer risk and preventive factor); and second, to measure the relationship between independent variables (breast cancer risk and preventive factor) and dependent variable (existence of cancer). Second, to continually learn and improve by exploring patterns and correlations in raw data over time and cases.

Remaining article structures as follows; Research methodology section to explain the search protocols and process of analysis followed by a discussion was done on the analysis results, followed by the conclusion.

RESEARCH METHODOLOGY

Literature Search for Analysis

To ensure replicability in future investigations, a systematic review method is used. The systematic review method is regarded as a trustworthy and scientific outline of existing body of knowledge in the field of research, to identify, review, and synthesize all relevant studies in a transparent and repeatable manner. The review process begins with the development of criteria and limitations for conducting a thorough literature search, as well as analysing and categorising the raw material ontologically. Precision, honesty, coverage, and full synthesis are all maintained.

Search protocols were established before the literature search for this research and were done in many stages. First, the search terms were finalized after the initial literature review, second, critical standards were set for the inclusion or exclusion of the articles from the searched one. Books, books chapter, and magazine articles were excluded from the searched materials, because of the unclear review process and limited diffusion. Journal articles and reports from the established and globally recognized cancer research institutes were included because of their validation as part of the body of knowledge. This method makes the results simple for reproduction.

These protocols brought a list of research articles and reports for the academic search engines i.e. ScienceDirect and PubMed. Every researcher then checked every study in full with the exclusion and inclusion criteria. This comprehensive practice finally results in 55 articles and reports discussing the application of AI and machine learning in cancer detection. On the final number of articles content analysis was performed, inductive approach was used to identify the application and benefit of AI and ML in early cancer detection. Procedures for search, exclusion and selection was as under;

A.) Criteria for selecting papers covering Application of AI & ML in cancer detection.

1. Cancer risk and preventive factor and detection through AI & ML
2. Articles of peer reviewed journal
3. Conceptual and empirical review and reports

Exclusion criteria on the basis of theoretical applicability

1. Studies in which primary focus is not application of AI & ML
2. Studies focused on cancer detection but not through AI & ML
3. Studies focused on cancer detection but not breast cancer
4. Research published in edited books and conference proceedings;
5. Research not electronically available or by other reasonable means.

Search Method and Scope - Stage–I

1. Full search of studies in academic journals related with the field
2. Sciencedirect and PubMed search engines
3. Inclusion scale by general keyword search using google scholar and google search engines.
4. Initial focus on abstract and title
5. Keywords;
6. Risk factors of breast cancer
7. Preventive factors of cancer
8. AI application for breast cancer detection
9. ML Algorithm for breast cancer detection
10. A focused search of key journals in the field to make sure that related articles not using keywords are included.

Search method and scope – Stage II

Manual evaluation of every article by experts, to include accept or except, to fit with set parameters of Stage-I.

Search method and scope – Stage III
Re-examine studies omitted by review but encompassed elsewhere (and added where believed as suitable)

Conducting Analysis
Themes emphasize the core concept of any research, therefore, the identified themes represent basic ideas, viewpoints, and theoretical connections of problem statements on which the research problem and assertions are based. Themes were extracted from the research articles on the rules of qualitative research coding. Objectives of the research, theoretical bases and methods helped to fetch primary and secondary themes. Themes were developed against the normal content analysis approach, which mines themes through de-contextualization. The process of theme identification and confirmation was extensive, iterative, and detailed. The first phase brought a substantial volume of themes, then sorted and categorized to organize according to ontological scope. The process of categorization reflects the context of the research. General convention of the ontological design was applied, with the establishment of superclass ascending to the subclass and alike, this was all done after arduous examination to avoid any repetition and redundancy at every level.

RESULTS AND DISCUSSION

According to the most recent statistics, breast cancer is rapidly increasing around the world, particularly in the Americas, Asia-Pacific, and Africa, prompting the requirement for early breast cancer detection to reduce mortality rates. A lot of research shows that early detection of cancer and its subtypes can lower death rates [4]. As a solution to the challenge of early cancer diagnosis, computer-aided programs with cancer-related experience were eventually developed. Early breast cancer diagnosis using machine learning (ML) for artificial intelligence (AI) is now expanding as a separate discipline within medical and computer science [6]. This was made possible by advances in statistics and artificial intelligence, which enabled medical practitioners, scientists, and AI experts to collaborate on prognostic development using factor analysis and regression analysis. The forecasts made by these hybrid systems are more accurate than those made by pure empirical predictions [23].

AI systems can filter out perplexing signals from large volumes of heterogeneous and multimodal data, progressively learning about the components involved in breast cancer and their effect [12]–[14]. AI has the ability to improve the capacity of the diagnosis system by commencing analysis in screened people based on clinically established parameters, thus, facilitating cancer detection [23].

Many machines and deep learning approaches have been developed over the last few decades to identify and classify breast cancer. These techniques may be divided into three categories: preprocessing, feature extraction, and classification. Preprocessing aids in the organization of data into a machine-readable format, as well as other associated requirements. The feature extraction procedure, on the other hand, identifies and distinguishes between malignant and benign tumors. [24]–[27].

Researchers commonly employ multi-modal neural networks to examine the difference between the values of the receiver operating characteristic (ROC) curve and the area under the curve by merging multi-dimensional data (AUC). The findings suggest that the effectiveness of breast cancer prediction may be enhanced by combining heterogeneous data with neural networks [23]. Another study amalgamates artificial neural networks with principal component analysis to find the patterns in the supplied data for the classification of the new cases. The study concluded that data and learning-based methods can help to develop an effective mechanism for prognostic studies by classifying cases into related categories more accurately depending on the severity of the tumor [28]. Other researchers have used other neural network versions to widen the use of neural networks and have produced more accurate results. With transcriptomics data, Cox-nnet (a neural network extension of the Cox regression model) predicts the occurrence, and the study has provided additional scientific knowledge [29]. Data from many modalities and dimensions, as well as classifiers and medication sensitivities, were used in the studies to employ a mix of artificial neural networks and principal component analysis to find patterns in data, which can be used to categorize fresh occurrences. According to the findings, data and learning-based techniques can help in the construction of an effective mechanism for prognostic research by more accurately classifying patients into suitable categories depending on the severity of the tumor [30], [31].

Random Forest Classifier (RFC) is another ML algorithm and is being used for the early detection of breast cancer. RFC belongs to the algorithm family that includes a sufficient number of decision trees that work as a group [32]. Every tree in the random forest makes a forecast and votes, and the tree with the most votes becomes the model's final prediction. The main benefit of RFC over other models and methodologies is that each tree (classifier) acts as a team member, and all members work together to arrive at a final forecast. Researchers discovered that RFC offers various advantages, including the ability to handle dichotomous classification, the ability to address datasets with fewer variables than observations, and the ability to work with heterogeneous data, all with high accuracy and efficiency [33]–[40]. RFC is being utilized for the prediction and classification of breast cancer, with different forms according to the field of application, because of its remarkable advantages and benefits.

The Support Vector Machine (SVM) is another frequently used ML algorithm. This algorithm traverses a hyperplane in an N-dimensional space, where N is the number of data-point classification features. The goal is to locate a plane with the greatest margin among the numerous different hyperplanes that might separate the two classes of data points. Margin refers to the data points in the two classes that are the furthest apart. Data points that fall on any side of a hyperplane can be categorized into distinct types. The hyperplane's dimension is determined by the number of features. SVs are locations that are closer to the hyperplane and have an impact on its orientation and position [20].

SVM is a seasoned and proven method for cancer detection and prognosis, and several studies have used it for various types of cancers; prostate cancer [41], Colorectal Cancer [42], lung cancer [43], Ovarian Cancer [44], Glioma [45], and Spinal Chordoma, Oral Cavity Squamous Cell, Pancreatic Neuroendocrine [46]–[48]. Researchers have also used SVM to estimate survival time for patients with breast cancer and compared it to other machine learning algorithms for accuracy and efficiency [49]. Another study employed SVM to diagnose breast cancer and came up with a hybrid algorithm for the best results [50]. SVM is a good classifier for clear margins, but it isn't suitable for large datasets since it takes longer to train and delivers disappointing results with noisy data.

In addition, to the above mentioned ML algorithms, there is another one known as logistic regression, which is a supervised machine learning method used by AI in and borrowed from statistics. Even though the name incorporates the word "regression," it is a classification model rather than a regression. It is a widely used categorization approach due to its ease of implementation and great results with linearly separable classes. A logistic regression model, like the Adaline and perceptron statistical approaches, classifies binary classes and may be expanded to multiclass classification [51]. The likelihood of a binary output, which might have any of the two values yes/no, true/false, and so on, is modeled using logistic regression. The logistic regression's multinomial variant can handle cases with more than two probable outcomes. It aids in determining the most appropriate categorization for a new sample [52].

The ease with which logistic regression may be implemented and the outstanding results it produces have led to its use in the identification and categorization of breast cancer. Several researches used logistic regression methods for machine learning and found that the results were good. For the identification of breast cancer, Seddik et al (2015) used the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and reported that findings derived via logistic regression outperformed alternative techniques[53]. Another study used logistic regression to train the computer to classify breast disorders as malignant or benign, concluding that the logistic regression model can clearly discriminate between the two types of tumors [54]. Another study used the logistic regression model to identify and classify breast cancer, with the findings being more accurate and exact [55].

This study mostly focused on the area of application of different ML algorithm, and their pros and cons. Furthermore, it only covers symbolic number of articles as the data and analysis, however, it is a first of its kind.

CONCLUSION

This study mostly focused on the area of application of different ML algorithm, and their pros and cons. Furthermore, it only covers symbolic number of articles as the data and analysis, however, it is a first of its kind, which used a qualitative approach for cancer research i.e. thematic analysis. Results of the study explored that machine learning algorithms are greatly helping the diagnosis of breast cancer at an early stage. There are two most widely used algorithms for this purpose with acceptable generalization capacity and clinical explainability; logistic regression, and neural networks. Logistic regression was used to measure the association between independent variables (breast cancer risk and preventive factor) and dependent variable (existence of cancer). Neural network algorithm applied to explore the patterns and correlations in raw data, over time and case and continuously learn and improve its intelligence.

As more sophisticated algorithms are being developed with more accurate data, more precisely machines are learning about the malignancies and their types without going through clinical and pathological procedures which are rather expensive or painful.

REFERENCES

[1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA.

Cancer J. Clin., vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] S. Chaudhury, M. Rakhra, N. Memon, K. Sau, and M. T. Ayana, "Breast Cancer Calcifications: Identification Using a Novel Segmentation Approach," Comput. Math. Methods Med., vol. 2021, pp. 1–13, Oct. 2021, doi: 10.1155/2021/9905808.

[3] A. Büyükavcu, Y. E. Albayrak, and N. Göker, "A fuzzy information-based approach for breast cancer risk factors assessment," Appl. Soft Comput., vol. 38, pp. 437–452, Jan. 2016, doi: 10.1016/j.asoc.2015.09.026.

[4] S. McPhail, S. Johnson, D. Greenberg, M. Peake, and B. Rous, "Stage at diagnosis and early mortality from cancer in England," Br. J. Cancer, vol. 112, no. S1, pp. S108–S115, Mar. 2015, doi: 10.1038/bjc.2015.49.

[5] L. Uttley, B. L. Whiteman, H. B. Woods, S. Harnan, S. T. Philips, and I. A. Cree, "Building the Evidence Base of Blood-Based Biomarkers for Early Detection of Cancer: A Rapid Systematic Mapping Review," EBioMedicine, vol. 10, pp. 164–173, Aug. 2016, doi: 10.1016/j.ebiom.2016.07.004.

[6] B. Hunter, S. Hindocha, and R. W. Lee, "The Role of Artificial Intelligence in Early Cancer Diagnosis," Cancers (Basel)., vol. 14, no. 6, p. 1524, Mar. 2022, doi: 10.3390/cancers14061524.

[7] P. Sasieni, "Evaluation of the UK breast screening programmes," Ann. Oncol., vol. 14, no. 8, pp. 1206–1208, Aug. 2003, doi: 10.1093/annonc/mdg325.

[8] R. Maroni et al., "ARTICLE A case-control study to evaluate the impact of the breast screening programme on mortality in England," Br. J. Cancer, vol. 2021, no. 124, pp. 736–743, 2021, doi: 10.1038/s41416-020-01163-2.

[9] L. J. Esserman, "The WISDOM Study: breaking the deadlock in the breast cancer screening debate," npj Breast Cancer, vol. 3, p. 34, 2017, doi: 10.1038/s41523-017-0035-5.

[10] Y. Xu, J. Kepner, and C. P. Tsokos, "Identify Attributable Variables and Interactions in Breast Cancer," J. Appl. Sci., vol. 11, no. 6, pp. 1033–1038, Mar. 2011, doi: 10.3923/jas.2011.1033.1038.

[11] Z. Li, S. Wang, and X. Lin, "Variable selection and estimation in generalized linear models with the seamless L0 penalty," Can. J. Stat., vol. 40, no. 4, pp. 745–769, Dec. 2012, doi: 10.1002/cjs.11165.

[12] K. Dembrower et al., "Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study," Lancet Digit. Heal., vol. 2, no. 9, pp. e468–e474, Sep. 2020, doi: 10.1016/S2589-7500(20)30185-0.

[13] S. M. Meystre, P. M. Heider, Y. Kim, D. B. Aruch, and C. D. Britten, "Automatic trial eligibility surveillance based on unstructured clinical data," Int. J. Med. Inform., vol. 129, pp. 13–19, Sep. 2019, doi: 10.1016/j.ijmedinf.2019.05.018.

[14] J. T. Beck et al., "Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center," JCO Clin. Cancer Informatics, no. 4, pp. 50–59, Sep. 2020, doi: 10.1200/CCI.19.00079.

[15] B. H. Kann, R. Thompson, J. Charles R. Thomas, A. Dicker, and S. Aneja, "Artificial Intelligence in Oncology: Current Applications and Future Directions," Oncology, vol. 33, no. 2, pp. 46–53, 2019, [Online]. Available: https://www.cancernetwork.com/view/artificial-intelligence-oncology-current-applications-and-future-directions

[16] W. W. Yim, M. Yetisgen, W. P. Harris, and W. K. Sharon, "Natural Language Processing in Oncology: A Review," JAMA Oncol., vol. 2, no. 6, pp. 797–804, Jun. 2016, doi: 10.1001/JAMAONCOL.2016.0213.

[17] S. Bhatia, Y. Sinha, and L. Goel, "Lung Cancer Detection: A Deep Learning Approach," in Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing, J. C. Bansal, Ed. Berlin, Germany, 2019, pp. 699–705. doi: 10.1007/978-981-13-1595-4_55.

[18] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," PLoS One, vol. 14, no. 3, p. e0214587, Mar. 2019, doi: 10.1371/journal.pone.0214587.

[19] A. R. Vaka, B. Soni, and S. R. K., "Breast cancer detection by leveraging Machine Learning," ICT Express, vol. 6, no. 4, pp. 320–324, Dec. 2020, doi: 10.1016/j.icte.2020.04.009.

[20] S. Bhise, S. Bepari, S. Gadekar, D. Kale, A. S. Gaur, and S. Aswale, "Breast Cancer Detection using Machine Learning Techniques," Int. J. Eng. Res. Technol., vol. 10, no. 7, 2021, [Online]. Available: http://www.ijert.

org

[21] Z. Baset, J. Abdul-Ghafar, Y. N. Parpio, and A. M. Haidary, “Risk factors of breast cancer among patients in a tertiary care hospitals in Afghanistan: a case control study,” BMC Cancer, vol. 21, no. 1, p. 71, Dec. 2021, doi: 10.1186/s12885-021-07798-5.

[22] C. Zangmo and M. Tiensuwan, “Application of logistic regression models to cancer patients: a case study of data from Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) in Bhutan,” J. Phys. Conf. Ser., vol. 1039, p. 012031, Jun. 2018, doi: 10.1088/1742-6596/1039/1/012031.

[23] S. Huang, J. Yang, S. Fong, and Q. Zhao, “Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges,” Cancer Lett., vol. 471, pp. 61–71, Feb. 2020, doi: 10.1016/j.canlet.2019.12.007.

[24] J. A. Cruz and D. S. Wishart, “Applications of Machine Learning in Cancer Prediction and Prognosis,” Cancer Inform., vol. 2, p. 117693510600200, Jan. 2006, doi: 10.1177/117693510600200030.

[25] M. F. Akay, “Support vector machines combined with feature selection for breast cancer diagnosis,” Expert Syst. Appl., vol. 36, no. 2, pp. 3240–3247, Mar. 2009, doi: 10.1016/j.eswa.2008.01.009.

[26] D. N. Ponraj, M. E. Jenifer, P. Poongodi, and J. S. Manoharan, “A survey of the preprocessing techniques of mammogram for the detection of breast cancer,” J. Emerg. Trends Comput. Inf. Sci., vol. 2, no. 12, pp. 656–664, 2011, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.651.592&rep=rep1&type=pdf

[27] G. Valvano et al., “Convolutional Neural Networks for the Segmentation of Microcalcification in Mammography Imaging,” J. Healthc. Eng., vol. 2019, pp. 1–9, Apr. 2019, doi: 10.1155/2019/9360941.

[28] S. Jhajharia, H. K. Varshney, S. Verma, and R. Kumar, “A neural network based breast cancer prognosis model with PCA processed features,” in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Sep. 2016, pp. 1896–1901. doi: 10.1109/ICACCI.2016.7732327.

[29] T. Ching, X. Zhu, and L. X. Garmire, “Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data,” PLOS Comput. Biol., vol. 14, no. 4, p. e1006076, Apr. 2018, doi: 10.1371/journal.pcbi.1006076.

[30] D. Sun, M. Wang, and A. Li, “A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 16, no. 3, pp. 841–850, May 2019, doi: 10.1109/TCBB.2018.2806438.

[31] A. Bomane, A. Gonçalves, and P. J. Ballester, “Paclitaxel Response Can Be Predicted With Interpretable Multi-Variate Classifiers Exploiting DNA-Methylation and miRNA Data,” Front. Genet., vol. 10, Oct. 2019, doi: 10.3389/fgene.2019.01041.

[32] L. Kuncheva, T. Christy, I. Pierce, … S. M.-C. on I., and U. 2011, “Multi-modal biometric emotion recognition using classifier ensembles,” in Springer, 2011, pp. 317–326. Accessed: Apr. 04, 2022. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-21822-4_32

[33] R. Díaz-Uriarte and S. Alvarez de Andrés, “Gene selection and classification of microarray data using random forest,” BMC Bioinformatics, vol. 7, Jan. 2006, doi: 10.1186/1471-2105-7-3.

[34] T. Khoshgoftaar, … M. G.-19th I. I., and U. 2007, “An empirical study of learning from imbalanced data using random forest,” in 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), 2007, pp. 310–317. Accessed: Apr. 04, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4410397/?casa_token=2G38dwBUA34AAAAA:9-lusLBducdwWRCDMk1OfH_beajk8lLD3-0okol8KcmGJWdQZQrmpKlGxBr0CwPxnslUaQjhPZc

[35] M. Khalilia, S. Chakraborty, and M. Popescu, “Predicting disease risks from highly imbalanced data using random forest,” BMC Med. Inform. Decis. Mak., vol. 11, no. 1, p. 51, Dec. 2011, doi: 10.1186/1472-6947-11-51.

[36] J. S. Evans, M. A. Murphy, Z. A. Holden, and S. A. Cushman, “Modeling Species Distribution and Change Using Random Forest,” in Predictive Species and Habitat Modeling in Landscape Ecology, New York, NY: Springer New York, 2011, pp. 139–159. doi: 10.1007/978-1-4419-7390-0_8.

[37] Z. Masetic and A. Subasi, “Congestive heart failure detection using random forest classifier,” Comput. Methods Programs Biomed., vol. 130, pp. 54–64, Jul. 2016, doi: 10.1016/j.cmpb.2016.03.020.

[38] M. Mohammady, H. R. Pourghasemi, and M. Amiri, “Land subsidence susceptibility assessment using random forest machine learning algorithm,” Environ. Earth Sci., vol. 78, no. 16, p. 503, Aug. 2019, doi: 10.1007/s12665-019-8518-3.

[39] A. Subasi, E. Alickovic, and J. Kevric, “Diagnosis of Chronic Kidney Disease by Using Random Forest,” in CMBEBIH 2017. IFMBE Proceedings, V. 62, Ed. Singapore: Springer Singapore, 2017, pp. 589–594. doi: 10.1007/978-981-10-4166-2_89.

[40] M Belgiu and L Drăguţ, “Random forest in remote sensing: A review of applications and future directions,” ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 24–31, Apr. 2016, Accessed: Apr. 04, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616000265?casa_token=v9qGvhQ48v4AAAAA:xAOeL_Ph-zlaPbCtbPMwmiIRH-LKWEZpIQ6kHLfHPF3kO9fKfHN8EGruOuKS49Xr4YgDpEagv0c

[41] S. Zhang et al., “Improvement in prediction of prostate cancer prognosis with somatic mutational signatures,” J. Cancer, vol. 8, no. 16, pp. 3261–3267, 2017, doi: 10.7150/jca.21261.

[42] L. Bottaci et al., “Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions,” Lancet, vol. 350, no. 9076, pp. 469–472, Aug. 1997, doi: 10.1016/S0140-6736(96)11196-X.

[43] C. M. Lynch et al., “Prediction of lung cancer patient survival via supervised machine learning classification techniques,” Int. J. Med. Inform., vol. 108, pp. 1–8, Dec. 2017, doi: 10.1016/j.ijmedinf.2017.09.013.

[44] T.-P. Lu et al., “Developing a Prognostic Gene Panel of Epithelial Ovarian Cancer Patients by a Machine Learning Model,” Cancers (Basel), vol. 11, no. 2, p. 270, Feb. 2019, doi: 10.3390/cancers11020270.

[45] C.-F. Lu et al., “Machine Learning–Based Radiomics for Molecular Subtyping of Gliomas,” Clin. Cancer Res., vol. 24, no. 18, pp. 4429–4436, Sep. 2018, doi: 10.1158/1078-0432.CCR-17-3445.

[46] A. V. Karhade et al., “Development of Machine Learning Algorithms for Prediction of 5-Year Spinal Chordoma Survival,” World Neurosurg., vol. 119, pp. e842–e847, Nov. 2018, doi: 10.1016/j.wneu.2018.07.276.

[47] C. Lu, J. S. Lewis, W. D. Dupont, W. D. Plummer, A. Janowczyk, and A. Madabhushi, “An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival,” Mod. Pathol., vol. 30, no. 12, pp. 1655–1665, Dec. 2017, doi: 10.1038/modpathol.2017.98.

[48] H. A. Haenssle et al., “Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” Ann. Oncol., vol. 29, no. 8, pp. 1836–1842, Aug. 2018, doi: 10.1093/annonc/mdy166.

[49] I. Mihaylov, M. Nisheva, and D. Vassilev, “Machine Learning Techniques for Survival Time Prediction in Breast Cancer,” 2018, pp. 186–194. doi: 10.1007/978-3-319-99344-7_17.

[50] H. Asri, H. Mousannif, and H. Al Moatassim, “A Hybrid Data Mining Classifier for Breast Cancer Prediction,” [51] S. Raschka, Python Machine Learning. Birmingham: Packt Publishing, 2015. [Online]. Available:

https://books.google.com.pk/books/about/Python_Machine_Learning.html?id=GOVOCwAAQBAJ&printsec=frontcover&source=kp_read_button&hl=en&redir_esc=y#v=onepage&q&f=false

[52] T. W. Edgar and D. O. Manz, “Exploratory Study,” in Research Methods for Cyber Security, Elsevier, 2017, pp. 393–404. doi: 10.1016/B978-0-12-805349-2.00030-3.

[53] A. F. Seddik and D. M. Shawky, “Logistic regression model for breast cancer automatic diagnosis,” in 2015 SAI Intelligent Systems Conference (IntelliSys), Nov. 2015, pp. 150–154. doi: 10.1109/IntelliSys.2015.7361138.

[54] J. Chhatwal, O. Alagoz, M. J. Lindstrom, C. E. Kahn, K. A. Shaffer, and E. S. Burnside, “A Logistic Regression Model Based on the National Mammography Database Format to Aid Breast Cancer Diagnosis,” Am. J. Roentgenol., vol. 192, no. 4, pp. 1117–1127, Apr. 2009, doi: 10.2214/AJR.07.3345.

[55] Z. Khandezamin, M. Naderan, and M. J. Rashti, “Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier,” J. Biomed. Inform., vol. 111, p. 103591, Nov. 2020, doi: 10.1016/j.jbi.2020.103591