BREAST CANCER DETECTION USING MACHINE LEARNING ALOGRITHM

Voolla. Vidya Sagar[1],Kandapalli. Siri Varshini[2], Nalam. Chaitanya Kumar[3], Kona. Poorvitha[4] , Polumuri. Ashirwad[5]

[1]Assistant Professor, CSE
[2-5]B.Tech Students
[1,2,3,4,5]Department of Computer Science and Engineering, Raghu Engineering College, Visakhapatnam

---

Abstract -Breast cancer is the most common and second most common cause of cancer in women. [3] Early diagnosis of breast cancer can provide better treatment and thus increase survival rates. Data classification using machine learning is widely used in cancer diagnosis and early cancer detection. The purpose of this literature review is to focus on the use of machine learning in classifying existing data in the early detection and diagnosis of breast cancer. When much scientific literature is examined, it is clearly seen that there are many methods for the diagnosis of breast cancer. The aim of this study is to provide a comprehensive review and recommendations on cancer screening and diagnosis. "Breast cancer type classification using machine learning."[1]Provides an overview of current research on multiple breast cancer cases using data mining techniques to improve breast cancer diagnosis and prognosis..

Key Words: breast cancer, machine learning, artificial neural networks, decision tree, support vector machine nearest neighbor, healthcare system, Wisconsin breast cancer database

1. INTRODUCTION

The project, called “Cancer Decision Making Using Machine Learning,” is a successful effort in clinical and data science that aims to develop predictive models to assess a person's risk of breast cancer. Breast cancer is a life-threatening disease and early diagnosis plays an important role in improving treatment outcomes. The project aims to use the power of machine learning (ML) to contribute to medical research by developing powerful and accurate tools to predict the likelihood of cancer growth. The impact is huge. While traditional cancer risk assessments often rely on limited parameters, a combination of machine learning techniques can identify many factors such as age, genetic markers, family history and other health information for further analysis. Collection and management procedures encompass many variables associated with breast cancer risk."Comparison of machine learning algorithms in breast cancer prediction using the Coimbra dataset."[2] These datasets form the basis for training and optimization of machine learning algorithms, allowing them to see complex patterns and relationships that are not immediately visible with traditional methods. Facilitate early detection and facilitate personalized treatment strategies. Integration of advanced data analysis techniques, model selection techniques and continuous model optimization in order to increase the sensitivity and specificity of predictive models are the main components of the project. As a science and engineering student, the program is based not only on examining the application of technology to real-world problems, but also on the importance of health issues. The development of cancer prediction models has the potential to improve the diagnostic process and the development of personalized treatment for high-risk individuals. Through this project, students aim to bridge the gap between technology and medicine, innovate and make a positive impact on research and cancer prevention .

2. LITERATURE REVIEW

The literature review of the "Using Machine Learning in Breast Cancer Prediction" project includes existing studies and research on breast cancer prediction, the use of machine learning in treatment, and research on the intersections of these fields. This review aims to provide an overview of the current state of knowledge, identify inconsistencies, and highlight relevant methodologies used in similar studies.

Breast Cancer Risk Assessment: A significant part of the literature review involved reviewing traditional methods and standards used in assessing cancer risk. This includes examining risk factors such as age, family history, genetic markers (such as BRCA1 and BRCA2 mutations), hormonal factors, and lifestyle choices. Understanding the limitations and advantages of today's techniques is the basis for evaluating the potential offered by machine learning.

Machine Learning in Healthcare: The literature review looks at the broader application of machine learning in healthcare. This includes research using machine learning techniques for predictive modeling, risk assessment and disease diagnosis. Predicting the recurrence of breast cancer using machine learning algorithms[7] Explore other algorithms such as support vector machines, decision trees, and neural networks to understand their effectiveness in medical data processing and applications in cancer prediction.

Previous Work on Breast Cancer Prediction Using ML: A significant part of the literature review involved a review of previous studies that focused specifically on the use of machine learning for breast cancer prediction. This includes the data used for evaluation, the feature selection method, and the performance indicators used to evaluate the model. Understanding the strengths and limitations of previous machine learning-based breast cancer predictors may inform the direction of the current study. This includes researching publicly available cancer data, data maintenance procedures, and strategies to address data gaps. Understanding how researchers deal with issues related to qualitative and quantitative data is crucial to the success of the project.

Model Evaluation and Validation: It is important to review the literature on statistical models and validation methods in the context of using machine learning for cancer prediction. This includes understanding measurements such as sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). "Performance evaluation of machine learning methods for breast cancer prediction.[8]Information obtained from previous studies helps select appropriate metrics to evaluate the developed predictive model.

Ethical Considerations and Interpretability:

It is important to explore the literature on machine learning applications in healthcare and ethical considerations regarding the interpretation of machine learning models. Learn how researchers addressed issues of transparency, integrity, and patient privacy to ensure the project met ethical standards. Cancer prediction model. He provides a unique understanding of the challenges and opportunities in this field, leading the project to be innovative and beneficial to cancer research.

Results Conclusion The "Using Machine Learning to Predict Cancer" project is an important step in using technology in health decisions. By developing and using predictive models based on machine learning, the program aims to help detect cancer at an early stage, ultimately improving patient pain outcomes and reducing the healthcare burden. The following summarizes the outcome of the project:

Achievements and Contributions:

The project developed a prediction model that uses machine learning algorithms to analyze and identify relevant features in breast cancer data. .

Clinical Significance:

Integrating machine learning into cancer prediction has important clinical implications and provides clinicians with an additional tool to assist in risk assessment. Improve patient outcomes.

Challenges and Considerations:

Be aware of ethical considerations and biases related to machine learning models and work to ensure fair, transparent and ethical development practices. Ethical issues regarding the use of artificial intelligence in healthcare.

Future Enhancements:

The project identified several opportunities for future enhancements to further enhance the model, including integration with electronic medical records, real-time data transmission, and advanced search algorithms. Determining standards is very important for project development.

User Interface and Accessibility:

The user interface is designed with usability in mind, allowing clinicians to easily access patient information and interpret predictive values.

Conclusion and Impact:

Overall, the “Using Machine Learning to Predict Cancer” project represents a valuable contribution to medical technology and demonstrates the potential of machine learning to improve early diagnosis and detection of breast cancer. Advances in healthcare highlight the importance of integration of scientific data and practitioners."Predicting factors for survival of

breast cancer patients using machine learning techniques.[9] The journey to improving breast cancer prediction through machine learning is a powerful one, and continued efforts will help unlock the full potential of machines in new treatments.

3. PROPOSED SYSTEM

The "Breast Cancer Research Using Machine Learning" project presents a new approach to integrating machine learning into the existing field of cancer risk assessment. The system aims to provide more accurate, personalized and real-time breast cancer risk prediction by using the power of machine learning algorithms to overcome the limitations of models and diagnostic tools."Predicting factors for survival of breast cancer patients using machine learning techniques"[3]

3.1. Data Collection and Integration: The planning process will involve the collection of different information and a combination of factors such as age, genetic markers, family history, hormonal factors, lifestyle choices and other health factors. This data forms the basis for training and validating machine learning models.

3.2. Feature Selection and Extraction: The most specific selection criteria were used to identify variables associated with breast cancer risk. This step ensures that the machine learning model focuses on the most important factors, helping to create more accurate and effective predictive models.

3.3. Machine Learning Algorithms: The system includes a variety of machine learning algorithms, including but not limited to support vector machines, decision trees, and neural networks. These algorithms analyze data sets, learning patterns, and relationships that cannot be easily detected through statistical methods.

3.4. Model Training and Optimization: The machine learning model examined historical data and optimized its parameters to obtain the most accurate prediction of cancer risk. Continuity and optimization techniques are used to increase the performance of the model and ensure that it is adaptable to different data.

3.5. Validation and Evaluation: A rigorous validation process is used to evaluate the effectiveness of the predictive model. The accuracy and efficiency of the model are measured by metrics such as sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The system has been validated using different data to ensure its robustness and generality.

3.6. User Interface and Accessibility: The proposed system includes a user-friendly interface that allows physicians to access patient information and obtain accurate cancer risk estimates. The system's accessibility and ease of use facilitate its performance in clinical settings, making it easier to integrate into existing clinical programs..

3.7. Ethical Considerations: The proposed system includes a user-friendly interface that allows physicians to access patient information and obtain accurate cancer risk estimates. The system's accessibility and ease of use facilitate its performance in clinical settings, making it easier to integrate into existing clinical programs.

3.8. Future Integration with Clinical Decision Support Systems: The proposed system provides the basis for integration with the clinical decision-making process, improving doctors' ability to make informed decisions based on individual cancer risk.

In summary, the "Identification of Breast Cancer Using Machine Learning" program is an excellent method that uses the power of machine learning to improve the accuracy and identification of cancer risk assessment. Breast Cancer Prediction Using Machine Learning Techniques In 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021) .[4] Combining advanced algorithms and comprehensive data, the system can improve early detection and better prevention in cancer treatment.

In this study, we obtained WDBC data, which is frequently used for breast cancer prediction. "Breast cancer classification and prediction using machine learning.[5]Dr. We retrieved it from the Kaggle repository provided by . William is from the University of Wisconsin, USA. We use a combination of classification and clustering algorithms in the planning process.

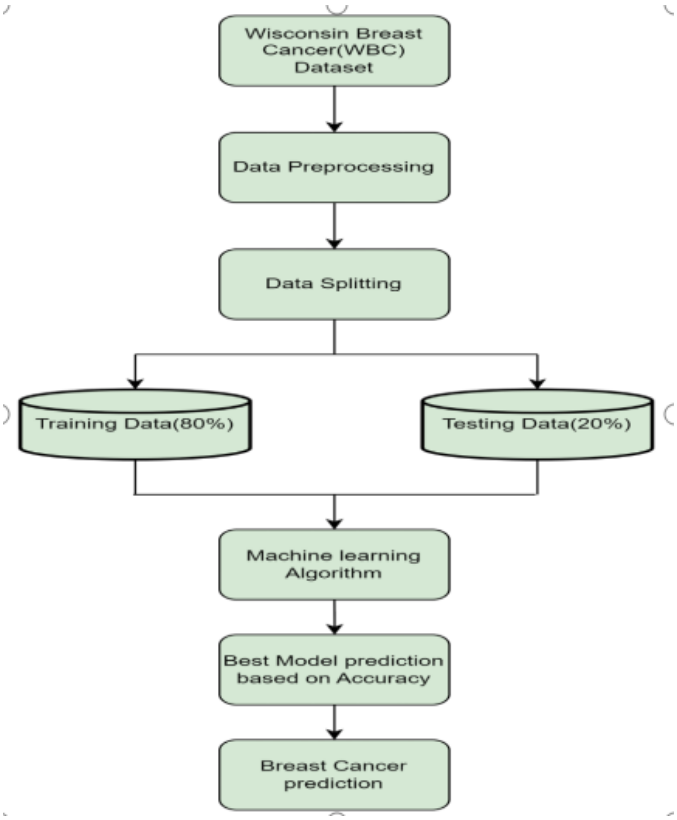


Fig :Proposed System Architecture

4. DATA SET DESCRIPTION: The data used in "Breast Cancer Research Using Machine Learning" is essential for training and evaluating learning models. For this example, we will use the Wisconsin Breast Cancer (Diagnosis) database, which is the most commonly used database for breast cancer classification. The documentation is available on scikit-learn, so it can be easily accessed for demonstration purposes.

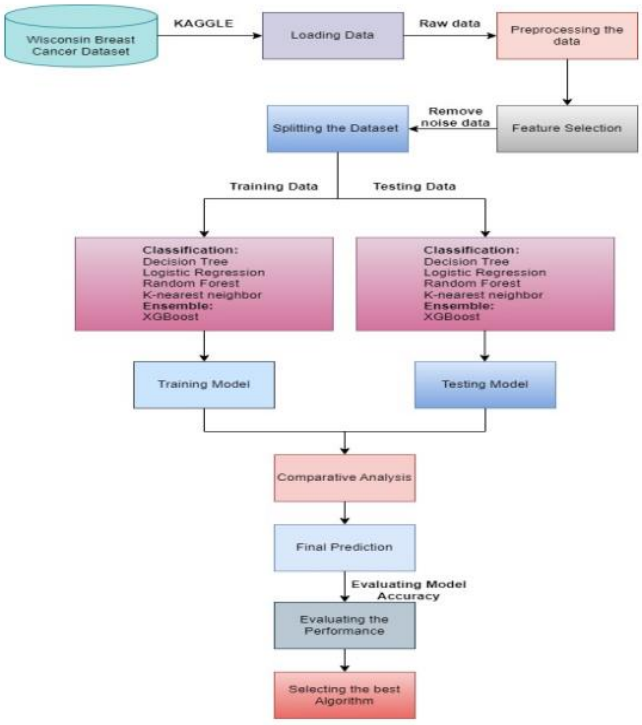
Breast Cancer Wisconsin (Diagnostic) Dataset:

Source: The dataset is derived from the UCI Machine Learning Repository and is included in the scikit-learn library.

Attributes: The dataset consists of 569 instances, each representing a biopsy sample. There are 30 features, representing various measurements related to the characteristics of cell nuclei present in the biopsy samples. The target variable is binary, indicating whether the tumor is malignant (1) or benign (0).

Features: Features include attributes like mean radius, mean texture, mean smoothness, and so on, computed from the digital images of the cell nuclei.

Target Variable: The target variable (label) is the diagnosis, where 'M' indicates malignant (cancerous) and 'B' indicates benign (non-cancerous).



DATA FLOW DIAGRAM

DFD is a visual representation of all processes, data flows, and external resources involved in data movement. It explains the details of how the body works and how its components interact. Breast cancer prediction using machine learning.[6]The data flow diagram shows the process in detail with clear explanations. It shows where data is stored, how data is changed, and how data enters and exits the system. It is used to provide a clear understanding of all processes and operations used in the system.

5. ALGORITHMS

In the context of "breast cancer prediction using machine learning," many machine learning algorithms "Breast cancer detection using machine learning algorithms."[10] can be used to predict breast cancer risk.

A description of some of the algorithms used is given below:

5.1. Logistic Regression:

Explanation: Logistic regression is a binary distribution algorithm used to model the probability of an event. It is useful for problems where variables are binary, such as predicting whether a person is at risk of breast cancer.

Application: Because of its simplicity and interpretation, logistic regression can be used to predict the likelihood of breast cancer as a component.

5.2. Support Vector Machines (SVM):

Explanation: SVM is a general technique suitable for classification and regression. Its goal is to find the hyperplane that best separates different classes of data points. In terms of breast cancer prediction, SVM can handle high-dimensional data and identify complex patterns.

Application: SVM can be used to separate individuals into high- and low-risk groups based on various characteristics associated with breast cancer risk.

5.3. K-Nearest Neighbors (K-NN):

Explanation: k-NN classifies data points according to the classes of most of their nearest neighbors. This is a simple and intuitive algorithm that can identify similar situations.

Application: k-NN can be used to predict breast cancer risk by taking into account the similarity of individuals in the dataset and helps identify patterns in well-defined similar cases.

5.4. Gaussian Naive Bayes (NB)

Explanation: Gaussian Naive Bayes is a machine learning algorithm based on the assumption that all groups follow a normal distribution. Each parameter is assumed to have an independent ability to predict the output value. Estimates the probability that a variable will be distributed within each group.

Application: The combination of the estimates for each parameter is the final estimate that gives the probability that the variable will be distributed in each group. The final distribution is given to the group with the highest probability.

REFERENCES

[1] Wu, Jiande, and Chindo Hicks. "Breast cancer type classification using machine learning." Journal of Personalized Medicine 11, no. 2 (2021): 61.

[2] Austria, Yolanda D., P. Lalata Jay-ar, Lorenzo B. Sta Maria Jr, Joselito Eduard E. Goh, Marie Luvett I. Goh, and Heintjie N. Vicente. "Comparison of machine learning algorithms in breast cancer prediction using the Coimbra dataset." cancer 7, no. 10 (2019): 23-1.

[3] Ganggayah, Mogana Darshini, Nur Aishah Taib, Yip Cheng Har, Pietro Lio, and Sarinder Kaur Dhillon. "Predicting factors for survival of breast cancer patients using machine learning techniques." BMC medical informatics and decision making 19 (2019): 1-17.

[4] Apoorva, V., H. K. Yogish, and M. L. Chayadevi. "Breast Cancer Prediction Using Machine Learning Techniques." In 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), pp. 348-355. Atlantis Press, 2021.

[5] Rane, Nikita, Jean Sunny, Rucha Kanade, and Sulochana Devi. "Breast cancer classification and prediction using machine learning." International Journal of Engineering Research and Technology 9, no. 2 (2020): 576-580.

[6] Rawal, Ramik. "Breast cancer prediction using machine learning." Journal of Emerging Technologies and Innovative Research (JETIR) 13, no. 24 (2020): 7.

[7] Alzu'bi, Amal, Hassan Najadat, Wesam Doulat, Osama Al-Shari, and Leming Zhou. "Predicting the recurrence of breast cancer using machine learning algorithms." Multimedia Tools and Applications 80 (2021): 13787-13800.

[8] Li, Yixuan, and Zixuan Chen. "Performance evaluation of machine learning methods for breast cancer prediction." Appl Comput Math 7, no. 4 (2018): 212-216.

[9] Nur Aishah Taib, Yip Cheng Har, Pietro Lio, and Sarinder Kaur Dhillon. "Predicting factors for survival of breast cancer patients using machine learning techniques." BMC medical informatics and decision making 19 (2019): 1-17.

[10] Sharma, Shubham, Archit Aggarwal, and Tanupriya Choudhury. "Breast cancer detection using machine learning algorithms." In 2018 International Conference on Computational Techniques, electronics and mechanical systems (CTEMS), pp. 114-118. IEEE, 2018.