

# Breast Cancer Prediction with Machine Learning Support Vectro Machine Technique

Uday Patil<sup>1</sup>, Rajeshwari Dhandare<sup>2</sup>, Sayali Bhole<sup>3</sup>, Prajwal Patil<sup>4</sup>, Ashish T. Bhole<sup>5</sup>

<sup>1, 2, 3, 4</sup>UG Engineering Students, <sup>5</sup>Associate Professor

Department of Computer Engineering, SSBT's College of Engineering and Technology, KBCNMU, Jalgaon, Maharashtra, India

*chaturuday1@gmail.com*

*Received on: 01 April, 2024*

*Revised on: 29 April, 2024*

*Published on: 01 May 2024*

**Abstract** – Breast cancer has become one of the leading causes of cancer death among women due to lifestyle changes and other factors. Detecting breast cancer takes a long time with manual diagnosis, even with available diagnostic systems. Statistics from the International Agency for Research on Cancer (IARC) reveal that breast cancer has overtaken lung cancer as the most commonly diagnosed cancer in women worldwide. The overall number of cancer diagnoses has nearly doubled from 10 million in 2000 to 19.3 million in 2020. Today, one in 5 people worldwide develop cancer during their lifetime. The increasing number of women facing breast cancer is a significant issue, challenging to diagnose in the early stage. A solution requires accurate, precise, and quick diagnosis, employing the best machine learning techniques for predictive results. The system aims to change the traditional diagnostic approach and provide analytical predictions based on user inputs and database figures.

**Keywords-** Breast cancer, Diagnosis, Analysis, Machine learning, Prediction.

## I -INTRODUCTION

**B**reast cancer, now the leading cause of cancer deaths in women, has globally surpassed lung cancer in diagnoses, as of December 2020 IARC data. Over the past two decades, cancer diagnoses doubled from 10 million in 2000 to 19.3 million in 2020, reflecting a concerning trend

of increasing cancer incidence worldwide. These alarming statistics, the urgency for effective prevention and treatment strategies has never been more apparent. With one in 5 people worldwide expected to develop cancer, the burden on healthcare systems is substantial, necessitating innovative approaches to address this public health challenge. As estimates predict a 50 percent increase in cancer cases by 2040, there is an imperative to invest in proactive measures to mitigate the impact of this global epidemic. In this context, leveraging advancements in machine learning and artificial intelligence presents a promising avenue for enhancing cancer prediction and diagnosis. By harnessing the power of machine-learning algorithms and sophisticated assessment metrics like confusion matrix, accuracy, and precision, researchers can develop robust predictive models and diagnostic tools tailored to the unique characteristics of breast cancer. Through this research focus, we aim to contribute to the ongoing efforts to combat breast cancer by harnessing the potential of cutting-edge technology to improve patient outcomes and alleviate the burden on healthcare systems.

## II- LITERATURE SURVEY

Gayathri B. M and Sumathi C. P in [1], The study compares Relevance Vector Machine (RVM) with other machine learning algorithms for breast cancer detection,

demonstrating that RVM performs well even with reduced variables. RVM's accuracy is comparable or better than other methods, making it a promising choice for breast cancer diagnosis.

RVM offers probabilistic predictions, low computational cost, and effective inference.

The study focuses on a specific dataset limiting the generalizability of RVM's performance across diverse datasets. Further exploration of RVM's robustness and scalability is needed.

Gupta Ankur etc. all in [2], The research proposes enhancing breast cancer prediction with a Convolutional Neural Network and edge detection, resulting in improved accuracy and efficiency.

Edge detection reduces computation time and memory use.

Success relies on high-quality, diverse training data. Clinical validation is essential before real-world application.

Bharathi A and Natarajan A. M. in [3], The research proposes a gene selection and classification approach for accurate cancer diagnosis using ANOVA and classifiers. The method demonstrates reduced gene complexity, and potential cost savings in cancer testing.

Identifying small subset of genes that are crucial for accurate classification.

Multiple steps and parameter tuning could make its implementation more complex than simpler classification techniques. The method may not work well for other types of cancer or diseases might be limited. U.A.P, R.P.D, S.L.B, P.P.P BREAST CANCER PREDICTION.

Islam M. M etc. all in [4], The study presents a comparative analysis of various machine learning techniques for breast cancer prediction. Through meticulous experimentation and evaluation, the authors demonstrate the efficacy of these techniques in predicting breast cancer. The findings reveal promising results, indicating the potential of machine learning models in assisting medical professionals with early diagnosis and treatment planning for breast cancer patients.

The paper provides a thorough comparison of different machine learning techniques, offering insights into their performance in breast cancer prediction. The conclusions are supported by empirical evidence obtained

through rigorous experimentation, enhancing the reliability of the study. The research underscores the clinical significance of machine learning in the realm of breast cancer diagnosis, highlighting its potential to improve patient outcomes and streamline healthcare processes. By elucidating the strengths and weaknesses of various machine learning approaches, the study contributes valuable knowledge to the interdisciplinary field of healthcare and artificial intelligence.

The study may have a restricted scope in terms of the datasets utilized or the specific machine learning algorithms investigated, potentially limiting the generalizability of the findings. The accessibility and quality of the datasets used for training and testing the machine learning models may influence the reproducibility and applicability of the results. Certain machine learning techniques examined in the study might be complex or resource-intensive to implement, posing challenges for practical deployment in real-world healthcare settings. The paper might not extensively address ethical considerations associated with the deployment of machine learning models in healthcare, such as patient privacy, algorithm bias, and interpretability of results.

Huang Min-Wei etc. all in [5], The study investigates the application of Support Vector Machine (SVM) and SVM ensembles for breast cancer prediction. Through their research, the authors highlight the effectiveness of these techniques in accurately predicting breast cancer outcomes. Their findings suggest the potential utility of SVM-based models in clinical settings for aiding in the diagnosis and prognosis of breast cancer.

The paper delves into the application of both SVM and ensemble methods, providing a comprehensive exploration of these techniques in breast cancer prediction. The study demonstrates the high predictive accuracy achieved by SVM and SVM ensembles, indicating their suitability for assisting healthcare professionals in making informed decisions regarding breast cancer diagnosis and treatment. By focusing on breast cancer prediction, the research directly addresses a critical healthcare concern, showcasing the practical relevance of machine learning in medical settings. SVM and ensemble methods are known for their scalability, allowing for the efficient processing of large datasets, which is crucial in healthcare applications where data volumes can be substantial.

SVM and ensemble models may involve intricate mathematical formulations and parameter tuning, which could pose challenges for non-experts in implementing and interpreting these techniques. The study's findings may be limited in generalizability due to factors such as the specific dataset used or the choice of evaluation metrics, potentially limiting the applicability of the results to broader contexts. SVM and ensemble models may require significant computational resources, including processing power and memory, which could be a barrier for adoption in resource-constrained environments. The paper primarily focuses on SVM and SVM ensembles without providing a comparative analysis with other machine learning approaches, potentially limiting insights into the relative performance of these techniques compared to alternative methods.

Yixuan Li and Chen Zixuan in [6], The study conducted by Li and Chen provides a comprehensive evaluation of various machine learning methods for breast cancer prediction. Through meticulous performance assessment, the authors offer insights into the efficacy of these methods in accurately predicting breast cancer outcomes. Their findings contribute valuable knowledge to the field of medical informatics, aiding in the development of more accurate and reliable predictive models for breast cancer diagnosis and prognosis.

The paper demonstrates methodological rigor in evaluating multiple Machine Learning methods, ensuring a thorough assessment of their performance in breast cancer prediction. The findings are presented clearly, allowing readers to grasp the comparative performance of different machine learning techniques with ease. By focusing on breast cancer prediction, the research directly addresses a critical healthcare concern, highlighting the practical relevance of machine learning in improving disease diagnosis and management. The study's findings hold the potential for translation into clinical practice, offering insights that may aid healthcare professionals in making more informed decisions regarding breast cancer diagnosis and treatment.

The study may have a limited scope in terms of the dataset used or the machine learning methods evaluated, potentially limiting the generalizability of the findings to broader contexts. The accessibility and quality of the dataset used for training and testing the machine learning models may influence the reproducibility and

applicability of the results. The paper may exhibit bias towards certain machine learning algorithms, potentially overlooking the potential benefits of alternative methods that were not included in the evaluation. The paper may lack a comprehensive evaluation of performance metrics, potentially overlooking important indicators of model performance beyond accuracy, such as sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).

Youness Khourdifi and Bahaj Mohamed in [7], The study conducted by Li and Chen provides a comprehensive evaluation of various machine learning methods for breast cancer prediction. Through meticulous performance assessment, the authors offer insights into the efficacy of these methods in accurately predicting breast cancer outcomes. Their findings contribute valuable knowledge to the field of medical informatics, aiding in the development of more accurate and reliable predictive models for breast cancer diagnosis and prognosis.

Practical Application: The paper addresses a practical application of machine learning in healthcare, focusing specifically on breast cancer prediction and classification, which is a pressing issue in the field of oncology. By considering multiple machine learning algorithms, the study offers a comprehensive assessment of different approaches, providing a nuanced understanding of their strengths and limitations in the context of breast cancer prediction. The research presented in the paper has direct relevance to real-world healthcare settings, potentially informing the development of decision support systems or diagnostic tools to aid clinicians in diagnosing and managing breast cancer cases. Being presented at a reputable international conference (ICECOCS) adds credibility to the study, indicating recognition within the academic community and peer review of the research findings.

As conference papers typically have length constraints, the level of detail provided in the paper may be limited, potentially hindering a comprehensive understanding of the methodologies employed and the results obtained. The availability and quality of the dataset used for training and testing the machine learning models may not be adequately discussed, raising questions about the generalizability and reproducibility of the findings. The paper may not extensively discuss the choice of evaluation metrics used to assess the performance of the machine

learning algorithms, which could impact the interpretation and comparison of results. Practical considerations such as computational resources required for implementing the machine learning algorithms or potential challenges in deploying them in clinical settings may not be thoroughly addressed.

Mohammed Naji Amine etc. all in [8], Naji et al. present a study that explores the application of machine learning algorithms for breast cancer prediction and diagnosis. Through their research, they provide insights into the effectiveness of these algorithms in accurately predicting and diagnosing breast cancer cases. The conclusions drawn from their analysis contribute to the growing body of knowledge in medical informatics, offering valuable guidance for researchers and healthcare professionals working towards improving breast cancer detection and patient outcomes.

The paper reflects the latest advancements and trends in machine learning applied to breast cancer prediction and diagnosis, ensuring relevance and timeliness of the findings. The study likely adopts a comprehensive approach by considering multiple Machine Learning algorithms, enabling a thorough evaluation of their performance and suitability for breast cancer prediction and diagnosis. The research addresses a pressing healthcare issue with direct relevance to real-world clinical practice, potentially informing the development of decision support systems or diagnostic tools to aid in breast cancer detection and management. Published in *Procedia Computer Science*, the paper likely underwent peer review, ensuring the quality and credibility of the research findings within the academic community.

As *Procedia Computer Science* typically imposes constraints on paper length, the level of detail provided in the paper may be limited, potentially hindering a comprehensive understanding of the methodologies employed and the results obtained. The paper may not extensively discuss the dataset used for training and testing the machine learning models, including considerations such as data size, quality, and representativeness, which are crucial for the generalizability and reproducibility of the findings. The choice and justification of evaluation metrics used to assess the performance of the machine learning algorithms may not be thoroughly discussed, impacting the interpretation and comparison of results. Practical considerations such as the computational

resources required for implementing the machine learning algorithms or potential challenges in deploying them in clinical settings may not be fully addressed.

Weigel Marion T., and Mitch Dowsett in [9], Weigel and Dowsett's review provides a comprehensive overview of current and emerging biomarkers in breast cancer, focusing on their prognostic and predictive value. The authors summarize the existing evidence and discuss the potential clinical implications of these biomarkers, highlighting their role in guiding treatment decisions and improving patient outcomes.

The paper covers a wide range of biomarkers, offering a thorough examination of both established and emerging markets in breast cancer. The authors base their conclusions on a comprehensive review of the literature, providing evidence-based insights into the prognostic and predictive value of various biomarkers. The review discusses the clinical implications of biomarker testing, emphasizing their utility in informing treatment strategies and personalized patient care. Despite being published in 2010, the review likely laid a foundation for subsequent research in the field, providing valuable insights into the state of biomarker research at the time.

The review may be subject to publication bias, as it relies on published literature which may not represent the full spectrum of research findings. The conclusions drawn from the review may not be universally applicable, as biomarker performance can vary depending on factors such as patient population and disease subtype. Given the rapidly evolving nature of biomarker research, some information presented in the review may have become outdated over time, necessitating updates to reflect the latest advancements in the field. The paper may be challenging for non-experts to understand due to the technical nature of biomarker research and terminology.

Sun Yi-Sheng etc. all in [10], Paper provides insights into the risk factors and preventive measures associated with breast cancer. The authors summarize the existing evidence on factors contributing to breast cancer development and discuss potential strategies for prevention.

The paper likely offers a comprehensive review of various risk factors and preventive measures for breast cancer, providing a holistic understanding of the topic. The authors likely base their conclusions on a thorough review of scientific literature, offering evidence-based insights

into breast cancer risk factors and prevention strategies. The review likely discusses practical implications for individuals and healthcare providers, highlighting actionable steps for reducing breast cancer risk and promoting preventive behaviors. Being published in the International Journal of Biological Sciences, the paper is likely accessible to a wide audience of researchers and healthcare professionals interested in breast cancer prevention.

The paper may lack detailed analysis or discussion on specific risk factors or preventive measures, potentially limiting the depth of understanding on certain aspects of breast cancer prevention. The review may be subject to publication bias, as it relies on published literature which may not encompass all relevant research findings or perspectives. Given the dynamic nature of breast cancer research, some information presented in the paper may have become outdated over time, necessitating updates to reflect the latest advancements in the field. The paper may be challenging for lay readers to understand due to the technical nature of the topic and scientific terminology used.

### III- METHODOLOGY

Support Vector Machine (SVM) is a supervised machine learning algorithm, it can be used for both classification or regression challenges. SVM algorithm finds the closest point of the lines from both the classes and points are called support vectors.

#### 01. Data Preparation: -

The step is dedicated to obtain a dataset consisting features of the breast cancer tumor and corresponding class labels (indicating whether the tumor is malignant or benign). Further it is divided into training and testing sets.

#### 02. Selecting the Kernel Function: -

Appropriate kernel function for SVM is to be chosen. Here the Radial Basis Function (RBF) kernel is used, it is known for its flexibility.

$$(K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2))$$

#### 03. Training the Model: -

The SVM classifier is trained using the dataset collected at the initial step. Formulate the optimization problem to find the hyperplane that separates the classes with the maximum margin.

Gradient descent and quadratic programming are used to solve the optimization problem as per requirement.

#### 04. Making Predictions: -

The model is trained, using the testing dataset the predictions are made. For each data point in the testing set, the decision function is calculated:

$$f(x) = \text{sign} (\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b)$$

If  $f(x)$  is positive, the data point belongs to the malignant class; if negative, it belongs to the benign class.

#### 05. Model Evaluation: -

The performance of the model is evaluated using criteria such as accuracy, precision, recall, F1-score, and ROC curve analysis. The model's performance is compared with other algorithms or models.

#### 06. Regularization and Tuning Parameters: -

The parameters  $C$  and  $\gamma$  are tuned to optimize the model's performance. Techniques such as grid search and cross-validation are used to find the best combination of parameters.

### IV -DESIGN

Systems Architecture is a generic discipline to handle objects (existing or to be created) called "systems", in a way that supports reasoning about the structural properties of these objects. The system architecture is the conceptual model that defines a system's structure, behavior, and views. An architecture description is a formal description and representation of a system. It provides a broad understanding of the portal. In the system, the architecture database provides functionality like getting information, selecting criteria, etc. to users. The system architecture is as mentioned in Fig. 1.

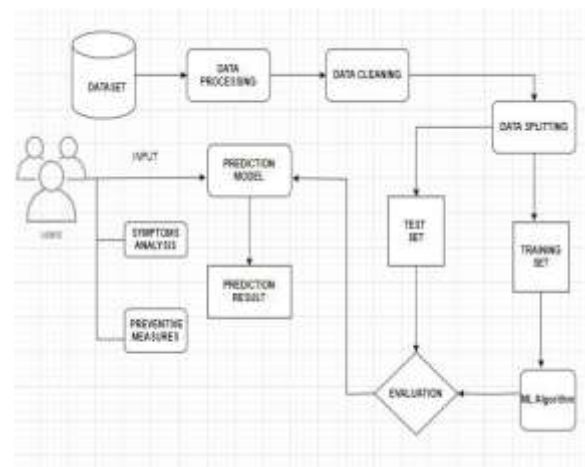


Fig. 1. System Architecture

Table 1 .Result

Sr. No.	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Result
1	8	10	10	8	7	10	9	7	1	Cancerous
2	8	7	5	10	7	9	5	5	4	Cancerous
3	10	5	5	3	6	7	7	10	1	Cancerous
4	5	1	1	1	2	1	3	1	1	Non-Cancerous
5	5	4	4	5	7	10	3	2	1	Non-Cancerous
6	3	1	1	1	2	2	3	1	1	Non-Cancerous

**LEVEL 0 DFD**

Level 0 contains one input and one output. The system provides information to the user means the system is the input and the user is the output. Fig. 2. shows the Level 0 DFD of the project.

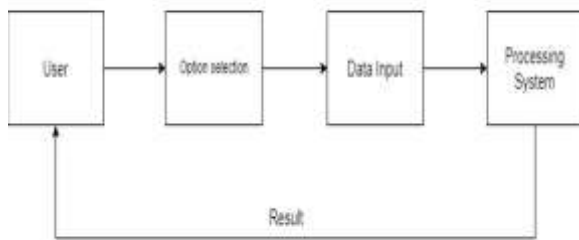
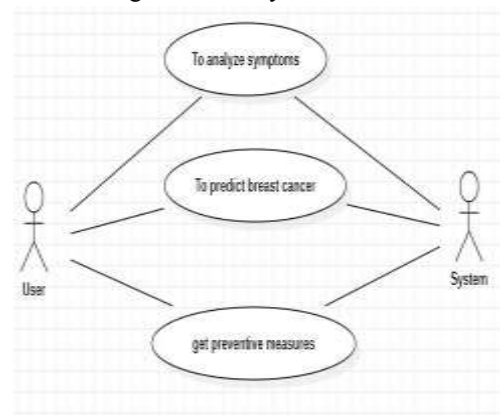


Fig. 2.- Level 0 DFD

**USE CASE DIAGRAM**

A Use Case Diagram is a visual representation in UML that depicts the interactions between actors (typically users or external systems) and a system or application. It provides a high-level view of the system’s functionality by illustrating various use cases, which are specific scenarios or actions that users or external entities can perform within the system. Actors are represented as stick figures, and use cases are displayed as ovals or ellipses connected by lines. Use Case Diagrams help in understanding the system’s functional requirements, identifying user roles, and

visualizing how users interact with the system, making them valuable tools for system design and communication between stakeholders. The Fig. 3. shows the Use Case Diagram of the system.



**V- RESULT**

Description about the main findings of a research, whereas the discussion section interprets the results for readers and provides the significance of the finding.

Support Vector Machine is ideal to use. To be tested SVM machine learning algorithm on the ‘Breast Cancer Dataset and reviewed the results and Breast Cancer Prediction using Support Vector Machine with and accuracy of 97.14. Results of Algorithm shown in Table 1. Results. Further is the line graph plotted for the results of accuracy of different algorithms. Fig. 4. Result Graph shows the analytical view.

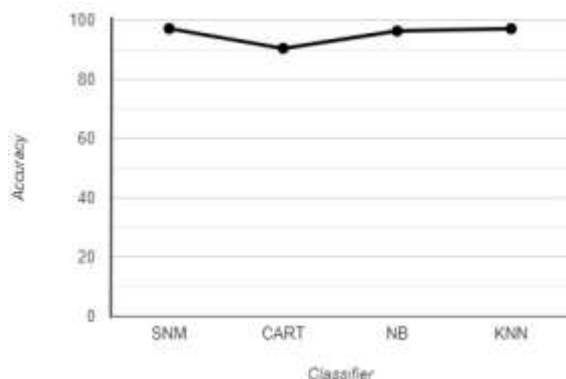


Fig. 4. Result Graph

Heading forward to prediction of the cancer using the true parameters, the result is printed as if the tumor is cancerous or not. Assuming some values, Table 2. Prediction, shows the input values and the respective output. The input parameters are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses, Result.

Table2- Prediction

Sr. No.	Name of classifier	Accuracy
1	Support Vector Machine (SVM)	97.14
2	Classification and Regression Tree (CART)	90.47
3	Naive Bayes (NB)	96.32
4	K-Nearest Neighbors (KNN)	97.10

## VI- CONCLUSION

In conclusion, the breast cancer prediction system utilizing machine learning demonstrates promising potential for accurate and early detection of breast cancer. By leveraging advanced algorithms and a comprehensive dataset, this system will be contributed to improved patient outcomes through timely diagnosis and treatment. As research and technology progress, integrating such systems into medical practice could lead to enhanced healthcare strategies and ultimately save lives. After implementation, the system went through testing phase, followed by the thorough observation of results to ensure the accuracy and

effectiveness of the breast cancer prediction system. The Breast Cancer Prediction System demonstrates promising potential for accurate and early detection of breast cancer. As future work, we propose enhancing the system's usability by creating user-friendly decision support tools and incorporating personalized risk profiles. We discuss the implications of integrating such systems into medical practice and their potential to enhance healthcare strategies and save lives.

For future enhancement one can create user-friendly decision support tools that empower patients and healthcare providers to make informed decisions about treatment and lifestyle changes based on personalized risk profiles.

## ACKNOWLEDGMENT

The success and final outcome of a project requires a lot of guidance and assistance from many people and it was extremely our privilege to have it all along the completion of project. All that we have done was only possible with such supervision and assistance and we would like to thank them and express our deep gratitude and sincere thanks to all who helped us to complete the project successfully. We would also like to thank Prof. Dr. Girish K. Patnaik, Principal for providing us facilities to complete the work. Deep gratitude goes to Dr. Manoj E. Patil, Associate Professor and Head of Department and Mr. Ashish T. Bhole, Associate Professor and Project Guide for granting us the opportunity to conduct the project work as well as his valuable suggestions and guidance at the time of need. Also a great thanks to Mr. Akash D. Waghmare, Assistant Professor and BE Project In charge for his valuable suggestions and guidance. Sincerely thankful and fortunate enough to get constant encouragement, support and guidance from all teaching staffs of Computer Department of SSBT's COET which helped us in successfully completing our project work. Also, would like to extend our sincere esteems to all staff in laboratory for their timely support.

## REFERENCES

- [1] Gayathri, B. M., and C. P. Sumathi. "Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer." In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1-5. IEEE, 2016.
- [2] Gupta, Ankur, Dushyant Kaushik, Muskan Garg, and Apurv Verma. Machine learning model for breast

- cancer prediction." In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 472-477. IEEE, 2020.
- [3] Bharathi, A., and A. M. Natarajan. Cancer classification using support vector machines and relevance vector machine based on analysis of variance features." *Journal of computer science* 7, no. 9 (2011): 1393.
- [4] Islam, Md Milon, Md Rezwatul Haque, Hasib Iqbal, Md Munirul Hasan, Mahmudul Hasan, and Muhammad Nomani Kabir. Breast cancer prediction: a comparative study using machine learning techniques." *SN Computer Science* 1 (2020): 1-14.
- [5] Huang, Min-Wei, Chih-Wen Chen, Wei-Chao Lin, Shih-Wen Ke, and Chih-Fong Tsai. SVM and SVM ensembles in breast cancer prediction." *PloS one* 12, no. 1 (2017): e0161501.
- [6] Li, Yixuan, and Zixuan Chen. Performance evaluation of machine learning methods for breast cancer prediction." *Appl Comput Math* 7, no. 4 (2018): 212-216.
- [7] Khourdifi, Youness, and Mohamed Bahaj. Applying best machine learning algorithms for breast cancer prediction and classification." In 2018 International conference on electronics, control, optimization and computer science (ICECOCS), pp. 1-5. IEEE, 2018.
- [8] Naji, Mohammed Amine, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, and Olivier Debauche. Machine learning algorithms for breast cancer prediction and diagnosis." *Procedia Computer Science* 191 (2021): 487-492.
- [9] Weigel, Marion T., and Mitch Dowsett. Current and emerging biomarkers in breast cancer: prognosis and prediction." *Endocrine-related cancer* 17, no. 4 (2010): R245-R262.
- [10] Sun, Yi-Sheng, Zhao Zhao, Zhang-Nv Yang, Fang Xu, Hang-Jing Lu, Zhi-Yong Zhu, Wen Shi, Jianmin Jiang, Ping-Ping Yao, and Han-Ping Zhu. Risk factors and preventions of breast cancer." *International journal of biological sciences* 13, no. 11 (2017): 1387