Frontend Website Implementation for Breast Cancer Classification System Using Machine Learning

1st Shania Wardani
School of Electrical Engineering
Telkom University
Bandung, Indonesia
shaniawardani@student.telkomuniversity.ac.id

2n Suryo Adhi Wibowo
School of Electrical Engineering
Telkom University
Bandung, Indonesia
suryoadhiwibowo@telkomuniversity.ac.id

3rd 3rd Koredianto Usman
School of Electrical Engineering
Telkom University
Bandung, Indonesia
korediantousman@telkomuniversity.ac.id

***Abstract* - Early detection of breast cancer is essential to improve patient survival rates. One way to be used for such detection is to develop a classification system based on genomic data, which can provide more accurate and efficient results. This study aims to design and implement a Streamlit-based website frontend, which functions as a breast cancer classification system interface using Machine Learning technology. This user interface is designed with ease of use and optimal user experience, allowing medical personnel to quickly access and understand the analysis results. The main features of this website include an educational dashboard about breast cancer, a simple and structured patient data input form, and predictive analysis results displayed in an interactive format and can be downloaded for further documentation purposes. Tests conducted on the front of this website show that the system response time to display the analysis results is no more than 5 minutes, making it an efficient solution in supporting medical decision-making. With an intuitive and easily accessible interface, this website makes it easy for medical personnel to perform breast cancer analysis faster and more accurately, supporting more effective early detection efforts.**

Keywords: Streamlit, User Interface, Breast Cancer, Website

I. INTRODUCTION

Breast cancer is a complex disease characterized by the abnormal growth and spread of cells in breast tissue, leading to the formation of tumors. If left untreated, these tumors have the potential to spread throughout the body and pose a serious threat. Typically, breast cancer begins in the milk ducts or glands of the breast. In its early stages, the condition is not immediately life-threatening, allowing for early detection. However, without intervention, cancer cells can infiltrate surrounding breast tissue, resulting in detectable lumps or thickening. Invasive breast cancer can progress to nearby lymph nodes or other organs through metastasis, which is the primary life-threatening aspect of the disease. According to the World Health Organization (WHO), in 2022, 2.3 million women were diagnosed with breast cancer, with 670,000 deaths reported globally. Breast cancer can affect women worldwide, regardless of age after puberty, with the incidence rate increasing with age. This issue is complex, particularly due to the challenge of accessible early detection methods for breast cancer [1].

Recent studies indicate that the mortality rate of breast cancer patients in Indonesia remains alarmingly high, standing at 70% in 2022, despite 40% of cases being detected at stages one and two [2]. The primary method used for genomic profile classification in Indonesia still relies on laboratory techniques like mammography classification. However, this traditional method takes around five days to classify genes, leading to delayed diagnosis and ultimately contributing to the high mortality rates associated with breast cancer in the country. This study aims to design a website application for the classification of genomic profiles associated with breast cancer. The website is intended to analyze and detect various types of breast cancer based on genomic profiles, providing in-depth classification analysis results for each case.

II. THEORY REVIEW

This research focuses on the website Breast Cancer Analysis Web. Users can utilize this website to detect breast cancer types, which will also provide high-accuracy results. Breast Cancer Analysis website is a solution that should be a quick step to classify the types of breast cancer. The website is built using Streamlit with the Python programming language, which allows for creating interactive and responsive user interfaces. For data management and backend, Supabase is used, which provides a scalable database solution that is easy to integrate with web applications. Streamlit facilitates real-time data processing and visualization, while Supabase ensures secure and efficient data storage and management. With this combination, the website can provide fast and accurate genomic analysis, support early detection of breast cancer, and provide the information needed for better medical decision-making.Theory review

A. Framework Streamlit

Streamlit is a popular open-source Python framework that allows for rapidly developing web applications for data science and machine learning. It has gained significant popularity due to its simplicity and the ability to integrate seamlessly with various data science libraries such as TensorFlow, Pandas, and Scikit-learn. Some key features of Streamlit include:

1. Real-time Interactivity: Streamlit enables the creation of interactive data applications that allow for dynamic updates without needing to refresh the page.
2. Ease of Use: With minimal coding and no requirement for frontend web development skills (e.g., HTML, CSS, JavaScript), it allows data scientists and machine learning engineers to deploy their models with little overhead rapidly.
3. Integration with Machine Learning Models: Streamlit provides built-in support for displaying model outputs, making it easy to showcase machine learning predictions, visualizations, and reports in real-time.

B. Machine Learning

Machine learning plays a crucial role in the automated classification of breast cancer. By training models on labelled data such as gene expression profiles and histological images, machine learning algorithms learn patterns to predict whether a patient's tumors is benign or malignant. This has been particularly useful in genomic studies, where the vast amount of data and complexity makes traditional methods inefficient. Supervised Learning Models: Algorithms such as Decision Trees and Artificial Neural Networks (ANN) are commonly used for breast cancer classification tasks. These models are trained on known data to predict new, unseen data. When using ANN, a model can be trained with 30 features, while a Deep Neural Network (DNN) can be trained with 6 features for enhanced feature representation.

1. Feature Selection and Engineering: The success of machine learning in this domain heavily depends on selecting the right features (e.g., gene expression levels, histological parameters) and preprocessing them effectively to improve model performance.

C. User Interface (UI) Design

The design of the UI is a critical aspect of making machine learning tools accessible to healthcare professionals. A well-designed UI ensures that the system is easy to navigate, quick to respond, and provides valuable insights clearly and understandably. Key principles of UI design include:

1. Minimalistic Design: Avoiding clutter and focusing on essential features and information.
2. Accessibility: Ensuring the system is usable by individuals with varying levels of technical expertise.
3. Responsiveness: The interface should perform well on different devices and respond quickly to user actions.

III. RESEARCH METHODS

A. Website Overview

The website is designed to integrate cloud computing, web development, and advanced computational models for efficient functionality. It features a user-friendly interface supported by secure authentication, cloud storage, and robust back-end processing. Additionally, the system incorporates machine learning and deep learning models to deliver accurate analytical results.

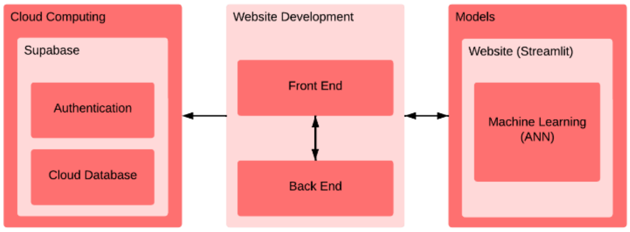

Figure 3.1
Website Implementation Diagram

Diagram 3.1 is a schematic showing how the chosen product system is implemented. Three subsystem implementation components comprise the system: cloud computing, website creation, and machine learning. The "Sequential" model is used by the TensorFlow and Keras frameworks in the learning (ANN) and (DNN) subsystems. Streamlit, a Python-based framework that enables the building of real-time prediction models with an interactive interface and direct interaction with deep learning and machine learning, is used in the website's design. Cloud computing systems use the SQL query language and JSON. This subsystem stores user information and analysis history using Cloud Supabase Database services. Streamlit is used to store models and provide website-building services in the deployment process.

The Website System Design Plan includes detailed activity diagrams to represent the key processes within the system. The Login Diagram Activity outlines the steps involved in user authentication, from input validation to data verification with the database, ensuring a secure and efficient login process in Figure 3.2. Additionally, the Analysis Diagram Activity illustrates the flow of data processing, where machine learning models analyze genomic profiles and synchronize results for user accessibility, as shown in Figure 3.3. These diagrams provide a comprehensive view of the system's design and functionality.

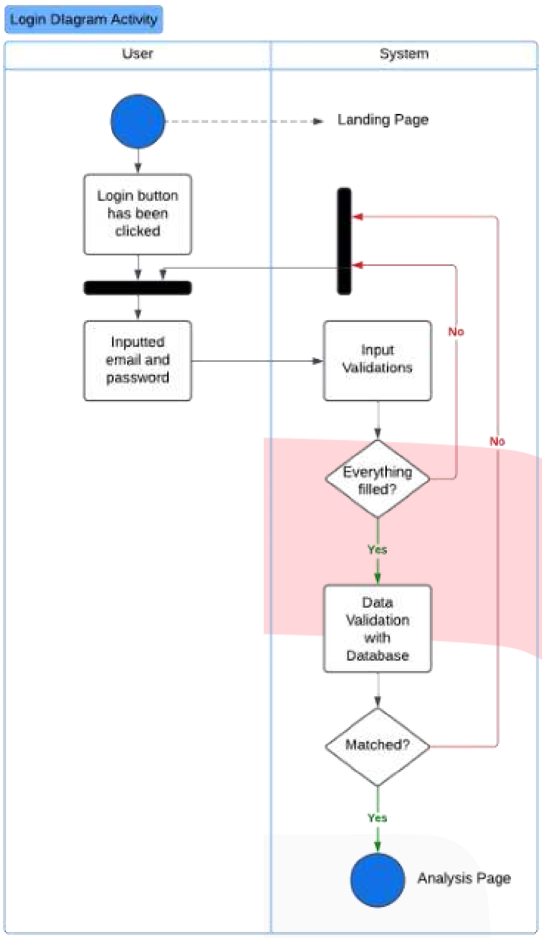

FIGURE 3.2
Login Diagram Activity

The activity diagram in Figure 3.2 illustrates the login process by dividing the flow into two main components: the User and the System. The process begins with the user clicking the "Login" button, after which they input their email and password. These credentials are then sent to the system for validation. Within the system, the first step is input validation, where the system checks if the email and password meet the required format and are non-empty. If this validation fails, the system sends an error message back to the user, prompting them to re-enter their credentials. Once the input passes the initial validation, the system proceeds to check whether all required fields are filled. If any field is incomplete, the user is notified to complete the form.

After ensuring completeness, the system conducts a data validation step by comparing the user's input with stored data in the database. If the provided credentials match an existing account, the login is deemed successful, and the user gains access to the system. Conversely, if the credentials do not match, an error message is returned, and the user is required to try again. The diagram uses red arrows to indicate failed processes, guiding the user to rectify errors, while green arrows signify successful steps that progress the process to the next stage. This structured flow ensures both security and user convenience, as every input is thoroughly validated before granting access.

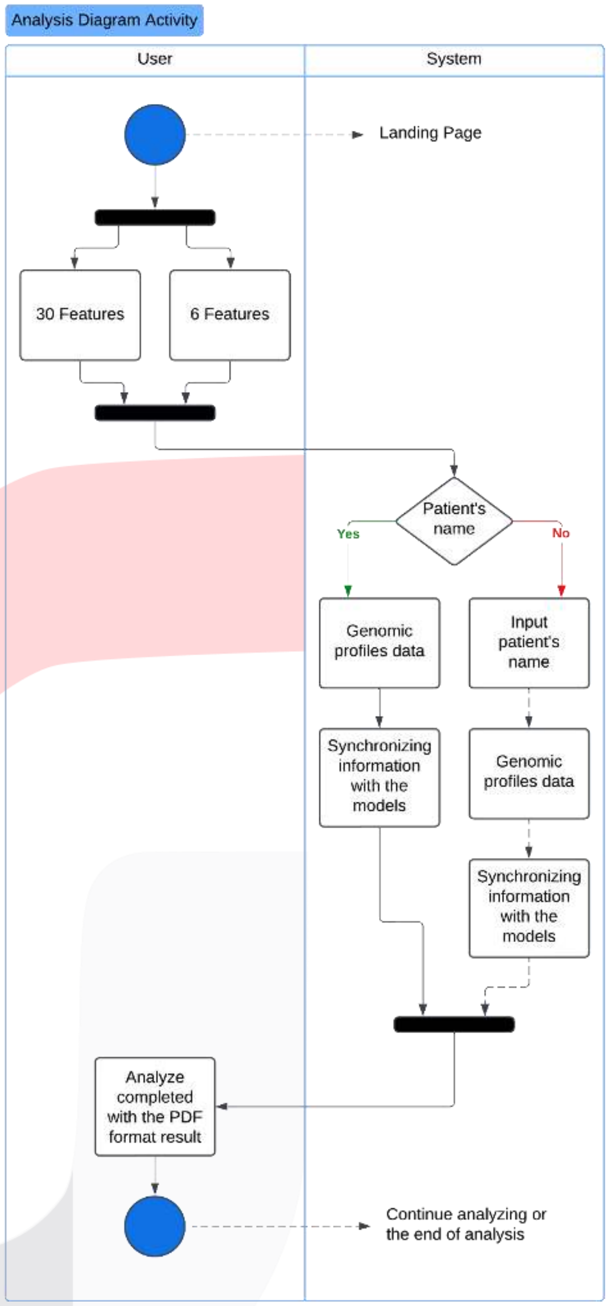

FIGURE 3.3
Analysis Diagram Activity

This activity diagram illustrates the analysis process, dividing the flow into two main components: the User and the System. The process begins with the user selecting ML Features (Machine Learning Features). Once the feature is selected, the flow proceeds with integrating the information into the system. In the System, the next step is to check for the presence of the patient's name. If the patient's name is available, the system retrieves the corresponding genomic profile data. However, if the patient's name is not available, the user is prompted to input the name, and the system then retrieves the genomic profile data. This ensures that the system has the necessary information to proceed with the analysis.

Once the genomic profile data is available, the system starts the process of synchronizing the information with the models. This step ensures that the data is prepared and aligned with the computational models used for the analysis.

After synchronization, the system completes the analysis and generates a result in PDF format, which is then provided to the user for review. The flow uses green arrows to indicate successful steps, while red arrows represent errors or missing data that require user input. This structured approach ensures data integrity and accuracy throughout the analysis process, making it efficient and user-friendly.

B. Frontend Architecture

The frontend system is designed to provide clear visualizations, intuitive navigation, and easy data input methods for medical professionals. The architecture includes the following components:

1. *Dashboard:* The dashboard serves as the main page, providing educational content and an overview of the system's capabilities. It includes informative sections on breast cancer prevention, treatment options, and the importance of early detection.
2. *Data Input Form:* This form allows users to input patient-specific data, such as gene expression levels and histological measurements. Input validation ensures that the data entered is consistent and within the expected range.
3. *Results Display:* The system presents predictive results from the machine learning model in an easy-to-understand format. The results include a classification label (benign or malignant), a confidence score, and a downloadable report.

C. Technologies Used

The development of the frontend system utilized several essential technologies to ensure functionality and user interactivity. Python was selected as the primary programming language due to its versatility and wide range of libraries. The Streamlit framework created an interactive and efficient user interface, simplifying the process of building and deploying the application. Matplotlib was used to generate static graphs for data visualization, while Plotly provided dynamic and interactive visualizations for enhanced user engagement. The system also incorporated Scikit-learn and TensorFlow to train and integrate machine learning models, enabling accurate data classification. The FPDF library was also implemented to generate downloadable PDF reports, allowing users to access and retain analysis results easily.

D. Website Testing

Testing is divided into two steps. The first step involves evaluating the usability of the Breast Cancer Analysis web application using the System Usability Scale (SUS). In this step, respondents interact with the application and provide feedback by completing a SUS questionnaire. The results are analyzed to measure the application's usability and ensure it meets user expectations for efficiency, effectiveness, and satisfaction. The second step involves testing the core functionality of the web application using the black-box testing method. This approach verifies that all features work as intended by testing inputs and outputs without examining the internal code structure. Testing is conducted on localhost to confirm that the system's functionalities meet their specified requirements.

IV. RESULT AND ANALYSIS

A. Implementation

Implementing an interface display, data retrieval, data upload, and visualization of analysis findings are some of the primary characteristics of the system. This website uses Streamlit, a Python framework that makes it simple to create interactive web pages for the interface display. Because of its effectiveness in creating interactive data applications, Streamlit has gained popularity [3]. Users may easily submit data and read prediction results in PDF format thanks to the executable web elements—such as input forms, buttons, and graphs—derived from the interface design previously produced in Figma.

The website loads the uploaded files for deep learning and machine learning using TensorFlow and Keras.H5 model formats that store the model and make predictions using the relevant input data as the model was included in the website. Currently, "joblib.load()" is used to get the previously saved files—"Label Encoder," "Standard Scaler," and "One-Hot-Encoder"—that were utilized in the data pretreatment procedure. Pandas and NumPy are used in the data processing part to manage and work with data that must be serialized into JSON format to be stored appropriately or transmitted over API [4]. The website connects to Supase as a backend service to store the prediction data in the database, enabling adequate data storage.

The Breast Cancer Analysis Website is developed using the Streamlit framework and includes four main features: Dashboard, Login, Sign-Up, and Analysis. The Dashboard provides users with an overview of key insights and visualizations. The Analysis feature allows users to input data, process it through integrated machine learning models, and view detailed classification results.

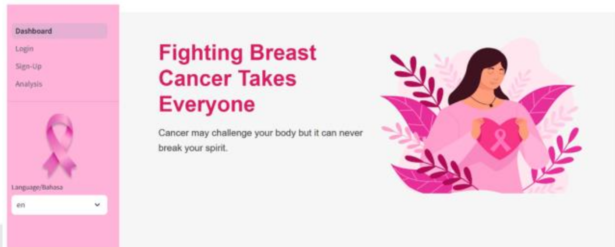

FIGURE 4.1
Dashboard Page

The Dashboard feature in the breast cancer analysis application is solely focused on breast cancer and application-related data. This section provides users with an explanation of breast cancer, including its signs, symptoms, and risk factors. It also discusses the methods that can be used to diagnose and treat the disease, as illustrated in Figure 4.1. The purpose of this article is to educate readers on the significance of early breast cancer detection and prevention strategies.

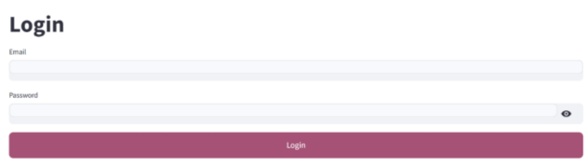

Figure 4.2 Login Page

The Breast Cancer Analysis website can be accessed by registered users by entering their login information in the Login section. For instance, while accessing the login section, the user is required to provide their already registered username and password. Once the login information is inputted, the data in the database is compared with the data entered by the system. After providing accurate login credentials, the user will be taken to the site's home page, which is shown in Figure 4.2. However, if the user tries to do the operation with the wrong set of data, an error message will show and they will be sent to another page. With these mechanisms in place, the website can tailor its offerings to the user's degree of comfort while limiting access to the information on the site to registered users and protecting the privacy of their data. This issue calls for a straightforward method that focuses solely on recognizing the necessity of the Login function.

Figure 4.3
Sign Up Page

The Breast Cancer Analysis website's Signup feature is shown in Figure 4.3. allows users to create a new account and register to enjoy the services offered. On the signup page, you must enter your complete name, email address, username, and password, among other personal information. The user will click the Register button after completing the registration form. The system will then examine the submitted data to make sure that no additional accounts (using the same email address or username) are registered. The user will click the Register button after completing the registration form. The system will then examine the submitted data to ensure that no additional accounts (using the same email address or username) are registered. When all the data entered is correct, a new account is created, and the user is taken to the login page, where they can complete the authentication procedures using the newly formed account. We have implemented this feature to allow new users to register and safely access the website. Furthermore, the Signup feature ensures that user information is legitimately registered and secure.

FIGURE 4.4
Analysis Page

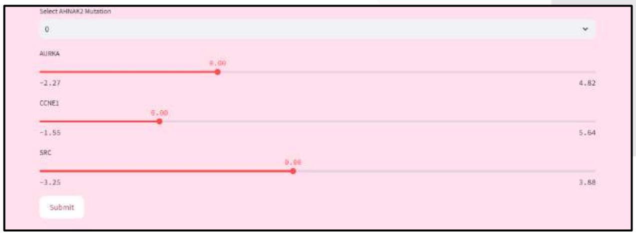

FIGURE 4.5
Column Input

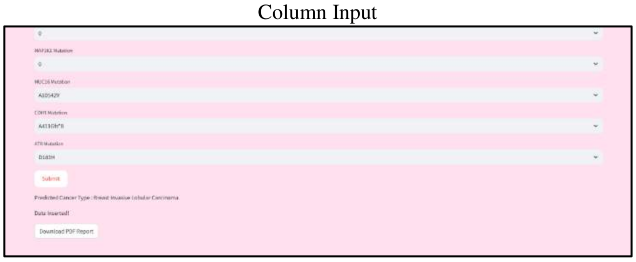

FIGURE 4.6
Submit and Prediction Button

The Breast Cancer Analysis website's Analysis section enables medical professionals to examine patient data entered using the given form, as illustrated in Figure 4.4, Figure 4.5 and, Figure 4.6. The user is prompted to enter medical information about the patient, including genetic mutations, histological data, and other essential factors for detecting breast cancer. The system uses a machine learning (ML) or deep learning (DL) based prediction model built on a medical dataset to process patient data after it has been entered automatically. Normalization and encoding of the input data are steps in the analysis process that ensure the data matches the model.

For example, StandardScaler will be used to scale the patient's numerical properties, and One-Hot-Encoder will be used to analyze categorical data, including the kind of genetic mutation. Following data processing, the machine learning model will use learned patterns to predict the potential cancer kind or cancer severity. The model's prediction results will be presented understandably, specifically as a prediction label that indicates the patient's potential breast cancer type. Additionally, this analysis result report includes recommendations for the required follow-up and visuals that aid in comprehending the prediction results. Medical staff can obtain the analysis report in PDF format upon completing the prediction process, which includes comprehensive details about the input data, predicted outcomes, and suggested treatments or preventative actions. In addition to giving medical professionals the information they need to take the appropriate medical action, this feature is anticipated to assist them in making quicker and more accurate judgments on how to manage patients with breast cancer.

B. Testing Method

The testing methods SUS (System Usability Scale) and Black Box Testing are essential for evaluating different aspects of a system. The System Usability Scale is a commonly utilized instrument for assessing the usability of a product, application, or system. The primary objective of SUS testing is to acquire a usability rating for the invention, which can then be re-evaluated and enhanced. This method can be successful for various products, including websites [5]. In SUS testing, a questionnaire is developed for data collection using platforms such as Google Forms. The questionnaire comprises ten items that evaluate several elements of the system, encompassing functional and usability metrics. All inquiries are evaluated using a five-point Likert scale ranging from Strongly Disagree to Strongly Agree. There are five affirmative and five negative inquiries, resulting in balanced and unbiased results. The questions are subsequently distributed to respondents, and their input is utilized to compute the SUS score. The results provide valuable insights into the system's strengths and areas for enhancement.

1. Testing Steps

The SUS was designed to assess satisfaction and, crucially, the usefulness of a use case or application. Participants are required to answer 10 questions using a Likert scale approach to assess the application's performance and usability. The following outlines the SUS test procedures and outcomes for the Breast Cancer Analysis website. Participants were recruited via convenience sampling, and data collection was conducted using the Google Forms platform, chosen for its accessibility and widespread utilization. This platform was selected to minimize complications for respondents while completing the survey. Ten questions were posed in the Google Forms survey to assess the system's usability and availability. The ten questions comprise five affirmative and five negative statements, intended to mitigate bias in respondents' quality assessments. The ultimate usability score was derived using the SUS methodology from the survey. The SUS scoring methodology is computed as follows: For affirmative questions, denoted by odd-numbered inquiries, the score contribution is calculated by deducting 1 from the selected scale value of the respondent. For negative questions, namely the even-numbered ones, the score contribution is determined by subtracting the respondent's selected value from 5. The entire score, after aggregating all contributions, is multiplied by 2.5 to yield the final usability score. Subsequently, it is employed to assess the website's usability [6]. the 10 survey questions are illustrated in Table 4.1.

TABLE 4.1
SUS Questions

No	Question
1	I feel that the appearance of this website is easy to understand.
2	I feel that this website is difficult to use to analysed breast cancer data.
3	I feel that the features on this website work according to their functions.
4	I feel that some of the features on this website are confusing.
5	I feel that the breast cancer analysis process on this website is quite simple.
6	I feel that I need a guidebook to use this website.
7	I feel that I do not experience any obstacles when using this website.
8	I feel that some things are inconsistent in the appearance or function of this website.
9	I feel that new users can understand how this website works easily.
10	I feel that I need to get used to it first before I can use this website smoothly.

Fifteen users of a website designed for breast cancer analysis participated in the survey. Each respondent was presented with a series of statements aimed at evaluating the usability and functionality of the website. These statements were rated using a Likert scale as illustrated in Table 4.2. which allows participants to express their level of agreement or disagreement. The scale included the following options:

Table 4.2 Likert Scale

Score	Description
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

2. Testing Result

This study involved 15 respondents using the Analysis Breast Cancer website. They provided feedback through a survey designed to evaluate the usability and functionality of the website. The collected data was analyzed using a specific formula to determine the website's overall usability score

$$x=((P1-1)+(5-P2)+(P1-3)+(5-P4)+(P5-1)+(5-P6)+(P7-1)+(5-P8)+(P9-1)+(5-P10))\times 2{,}5$$

Description:

x = Total value of each respondent

P = Value of each question

After obtaining the scores from each respondent, the next step is to calculate the average score from all respondents using the following formula:

$$\bar{x}=(\sum x)/n$$

Description:

$\bar{x}$ = Average score of respondents

n = Number of respondents

After all respondents completed the provided Google Forms and the calculations were performed using the formula above, the average usability score of the Breast Cancer Analysis website was 84.67, which corresponds to a Grade A. The detailed results can be found in Appendix A.

SUS Analysis Testing Result

Data collection in this study was done through a questionnaire of 15 respondents. The website's average score was obtained in usability. The mean score is obtained by summing the total scores of all respondents and dividing the sum by the number of respondents. After calculating the usability score for each respondent, an average score of 84.67 was obtained. The SUS score is then analyzed and interpreted using the acceptability category, grade scale, and adjective rating with a 10-point scale. In addition to these three methods, SUS provides another way to determine the assessment results through the SUS score percentile range, shown in Table 4.3.

Grade	Score

A	$x > 80{,}3$
B	$74 \leq x < 80{,}3$
C	$68 \leq x < 74$
D	$51 \leq x < 68$
E	$x < 51$

TABLE 4.3
Percentile Rannge

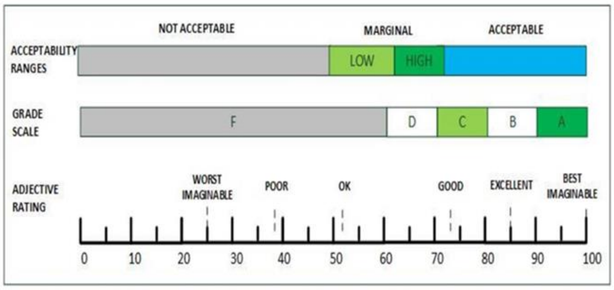

FIGURE 4.4
SUS Scores Comparison

Based on the test results of Figure 4.4. involving 15 respondents, an average value of 84.67 was obtained. In the first test method, which determines the Grade based on the Percentile Range, 84.67 is in the Grade A category, according to Table 5. 4. Meanwhile, in the second test method, 84.67 is included in the Acceptable category in the Acceptability Ranges. If adjusted to the Grade Scale, the value is included in Grade B because it is in the range of 80-90. In Adjective Ratings, 84.67 is categorized as Excellent because its value exceeds 80. From the results of this test, it can be concluded that the Breast Cancer Analysis website has good usability, is informative, and is relatively easy to use. However, further development is still needed, especially in improving features and display design, so the website's usability score can increase.

While Black box testing is an instance of software testing that relies on input and output without revealing the internal workings of the program or its code [7]. The purpose of this testing is to make sure the program or application is operating in accordance with the end user's requirements and expectations. Testing is carried out in the context of this study by assessing every feature and function on the Breast Cancer Analysis website. To make sure the outcomes are what is anticipated, each feature is evaluated using a predefined usage scenario.

3. Step of Testing

Black box testing uses scenarios that the tester creates to make that all of the website's functionalities function as intended. The tester needs to prepare the expected results before beginning the exam. Comparing the actual results with the anticipated results is how the testing is carried out. The test is classified as "Successful" if the findings meet the expectations. However, if the outcomes fall short of expectations, the test is labeled "Unsuccessful." The steps in the Black Box testing on the Breast Cancer Analysis website are as follows:
1. Create the testing scenarios using the requirements analysis for the website. The scenarios ought to encompass a range of possible circumstances that may arise throughout the testing procedure.
2. Provide specified inputs for each component of the Breast Cancer Analysis website to run the test in accordance with the scenarios that have been established.
3. Confirm that the problems found during testing have been fixed and that, following the implementation of the modifications, the website operates as intended.

The results of testing every page and feature on the Breast Cancer Analysis website demonstrated that everything worked as it should and produced the desired results. No Page/Feature Test Outcome Anticipated.

V. CONCLUSION

The development and implementation of the Breast Cancer Analysis website demonstrate an innovative approach to improving early detection and diagnosis using machine learning and intuitive web interfaces. Utilizing the Streamlit framework for its frontend architecture and a robust backend powered by Supabase, the system effectively integrates educational content, patient data analysis, and predictive tools. Usability testing, through the System Usability Scale (SUS), yielded a high score of 84.67, reflecting excellent performance, user satisfaction, and the system's overall functionality.

The website's key features, such as the dashboard, input forms, and analysis reporting, cater to the needs of medical professionals, enabling faster and more accurate decision-making. Black box testing confirmed that all functionalities perform as intended, further validating the system's reliability. While the current implementation has proven effective, further enhancements in feature development and interface design are recommended to elevate the system's usability and user experience. Ultimately, this project sets a solid foundation for leveraging technology in critical healthcare applications and highlights its potential to improve patient outcomes through early detection of breast cancer.

REFERENCES

[1] W. H. O. (WHO), "Breast Cancer," pp. https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 13 Maret 2024.

[2] M. d. Widyawati, "Kementerian Kesehatan," 2 Febuary 2022. [Online]. Available: https://www.kemkes.go.id/id/rilis-kesehatan/kanker-payudaya-paling-banyak-di-indonesia-kemenkes-targetkan-pemerataan-layanan-kesehatan.

[3] J. L. R. K. S. Smith, "Interactive Web Applications with Streamlit for Data Science," *Journal of Computational Science, 37(1), 45-58.*, 2023.

[4] C. R. M. K. J. v. d. W. S. J. G. R. &. V. P. Harris, "Array Programming with NumPy," *Computing in Science & Engineering (CSE),* vol. 13, no. 2, 2011.

[5] S. R. A. M. D. Moch Kosim, "PENGUJIAN USABILITY APLIKASI PEDULILINDUNGI DENGAN METODE SYSTEM USABILITY

SCALE (SUS)," *Jurnal Sistem Informasi Sains dan Teknologi,* vol. 4, no. 2, 2022.

[6] E. C. S. D. M. A. R. B. Endah Sumiati, "Pengujian Menggunakan Black Box Testing dengan Teknik State Transition Testing Pada Perpustakaan Yayasan Pendidikan Islam Pakualam Berbasis Web," *Jurnal Kreativitas Mahasiswa Informatika,* vol. 4, no. 2, 2022.

[7] J. Brooke, "SUS: A quick and dirty usability scale," *Usability evaluation in industry,* vol. 189, no. 194, pp. 4-7, 1996.