SLAS Technology 32 (2025) 100285

Contents lists available at ScienceDirect

SLAS Technology

journal homepage: www.elsevier.com/locate/slast

Full Length Article

Large language models in breast cancer reconstruction: A framework for patient-specific recovery and predictive insights

Chunrao Zheng *, Qunfang Li, Geling Lu, Yuchang Mai, Yuan Hu

Division of Breast Surgery, Department of General Surgery, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen 518020, Guangdong, PR China

ARTICLE INFO

Keywords:
Breast cancer reconstruction
Natural language processing (NLP)
Patient recovery
Large language models (LLMS)
Personalizing care

ABSTRACT

Breast cancer reconstruction, a vital part of comprehensive cancer therapy, can be performed concurrently with cancer resection, improving both physical and psychological recovery for patients. However, the intricacy and variety of recovery demand a specialized strategy. Thus, a unique framework that uses Natural Language Processing (NLP) and Large Language Models (LLMs) is developed to improve patient-specific recovery and predictive insights during breast cancer reconstruction. Lemmatization/Stemming is used for pre-processing large volumes of data from medical records, clinical notes, and treatment histories and BioBERT, a model pretrained on biomedical texts to capture complex medical terminology used for feature extraction and aids in the transformation of text data into numerical vectors. The approach employs forecasting models like ChatGPT-4 and Gemini to offer insights into the likelihood of successful reconstruction and associated problems based on specific patient characteristics, treatment options, and recovery timelines. Using sophisticated LLMs, this framework provides clinicians with a powerful tool for personalizing care by anticipating postoperative complications, recovery durations, and psychosocial consequences. Furthermore, it allows for the development of targeted rehabilitation programs that are adapted to unique patient needs, enabling greater recovery and overall quality of life. This approach not only improves clinical decision-making but also empowers patients by offering personalized recovery strategies. As a result, the accuracy of ChatGPT-4 is 98.4 % and Gemini is 98.7 %; the score per response is 2.52 for ChatGPT-4 and 2.89 for Gemini. Readability of ChatGPT-4 is 93.0 % and Gemini is 94.5 %; a relevance score is 95.5 % and 94.0 % for ChatGPT-4 and Gemini, and time response is 2.5 s for ChatGPT-4 and 2.5 s for Gemini. Finally, this research indicates how NLP and LLMs can transform breast cancer reconstruction by offering predictive insights and promoting tailored, patient-centered therapy, bridging the gap between powerful computational technologies and life science research to better patient care.

1. Introduction

The prevalence of breast cancer is increasing globally. Breast cancer is the most dominant type of cancer affecting millions of cases worldwide, surpassing lung cancer. Breast cancer ranks as the fourth most public cause of cancer-linked deaths worldwide which is the major cause of death among women [1]. The reduction in breast cancer mortality can be achieved by the improvements in medical technology and the increased use of breast cancer screening methods. It nonetheless showed that breast cancer treatment expenditures would reach almost billions [2]. There are numerous limitations in traditional methods, including the fact that they are single-center studies conducted by one surgeon and that surgical practice evolves with time and increased experience [3]. LLMs, such as Transformer-based Language Models (TLM), determineabilities in understanding and generating text like humanscreated thusenabling real-time statements and providing appreciated insights to health care authorities [4].

The swift advancement of Natural Language Processing (NLP) increases the medical field's curiosity about employing these methods to enhance the precision of cancer screening [5]. Recently, the launch of ChatGPT (OpenAI) provides tailored training feedback to surgical trainees and supports research in breast reconstruction [6]. LLMs are nonparametric models that utilize self-supervised learning and are trained on an extensive corpus of text data. They usually consist of a

* Corresponding author.
E-mail addresses: zcrao2011@163.com (C. Zheng), RhiannaQ7472@163.com (Q. Li), karelgl@163.com (G. Lu), m13509693025@126.com (Y. Mai), 15815571318@163.com (Y. Hu).

https://doi.org/10.1016/j.slast.2025.100285
Received 31 January 2025; Received in revised form 11 March 2025; Accepted 9 April 2025
Available online 10 April 2025

substantial number of parameters, spanning from hundreds of millions to billions. Nonetheless, with the right prompts, LLMs can achieve remarkable performance on a variety of NLP tasks [7]. In recent years, LLMs have progressed significantly. Recent advancements in computing power and model building have made these advanced models, which impose billions of constraints, possible.The potential LLMs might unlock better care for the patient through features such as assisted diagnosis, hasten medical science research through huge-scale literature researches, and improve the efficiency of the health system through automation [8]. The breast reconstruction can be performed either during the initial operation or following main surgery and oncological interventions. Further, surgical recommendations are provided by these LLM models for effective recovery [9]. However, there is a need for personalized recovery insights and predictive analytics in breast cancer reconstruction. Hence, enhanced care tailored to the individual patient, and improved treatment planningare essential for better recovery of patients.

1.1. Research objective

This research aims to enhance patient-specific recovery outcomes following breast cancer reconstruction by developing a predictive framework that utilizes NLP and LLM. It includes ChatGPT-4 and Gemini model that would predict recovery success, postoperative issues, and psychosocial effects by processing and correlating immense volumes of medical data, such as clinical notes and treatment history, to offer tailored insights to therapists. This method employs recent computational technology to make available patient-specific recovery programs, facilitate optimal decision-making, design individualized rehabilitation programs, and improve treatment in general.

1.2. Key contributions

- ➢ Patient data gathered comprises of medical records, treatment history, and clinical notes.
- ➢ Data pre-processing involves Lemmatization, stemming, and feature extraction using BioBERT.
- ➢ ChatGPT-4 and Gemini are used to predict recovery success and complications.
- ➢ Predictions for recovery duration, and rehabilitation programs to improve recovery strategies.
- ➢ NLP and LLM models aim to improve recovery outcomes based on accuracy, score per response, readability, relevance score, and response time.

1.3. Research organization

The following is the sequence in which the research is structured: Section 2 explains related works involving NLP and LLM models. A detailed summary of the systems utilized in the recommended method is given in Section 3. Furthermore, Section 4 offers the outcomes of the ChatGPT-4 and Gemini. Results of implemented models and limitations before the model's implementation are conferred in Section 5. Finally, the research's conclusion, and future workare determined in Section 6.

2. Literature review

The reduction of incidence of post-mastectomy lymphedema (PML) after axillary lymphadenectomy (AL) was performed by implementing a randomized Clinical Experiment (RCE) on the efficacy of theprimary lymphatic repair (PLR) method [10]. PLR reduced PML incidence (9.5 % vs. 32 %, $p = 0.014$) and improved lymphatic function, bioimpedance, andcompression use.The blinding and the use of an arbitrary 10 % RVC cutoff for diagnosing lymphedema should be improved. The assessment of decisional conflict in patients weighing immediate breast reconstruction revealed that 68 % faced clinically significant decisional conflict, especially among those with a lower reconstruction preference and greater anxiety [11]. Multivariable logistic regression analysis was employed, which producedrestrictions such as a sample composed of individuals with high levels of education and the possibility of overestimating decisional conflict.The effect of radiation dose on complications in breast cancer patients were analyzed [12]. In addition, major post-radiation therapy complications were found in 7.6 % of patients. The dependence on reviews of medical charts and differences in follow-up durations among institutions should be reduced.

Assessment on surgical outcomes, complications, regional recurrence, and reserved metastases in patients was analyzed who received autologous fat attaching following breast cancer surgery [13]. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) procedures were established, revealing a 1.1 % rate of local recurrence, a 2.2 % incidence of distant metastases, a 1.1 % tumor-related mortality rate, and a 67.12 % patient satisfaction level regarding autologous fat grafting. The research lackedreconsidering setting and the absence of a controller group.The depiction on lipofilling within breast reconstruction, patient satisfaction, and an examination of existing literature were performed [14]. The method was described in a multi-center retrospective cohort analysis conducted from 2013 January to 2020 of April. Data from 37 patients indicated an 18.9 % complication rate and a 5.4 % recurrence rate. The retrospective nature of the research showed that some biases cannot be excluded.The clinical facets and determinants linked to local recurrence (LR) following immediate breast reconstruction (IBR) were elucidated [15]. Invasive cancers had a 7-year LR rate of 4.3 %, with survival rates in LR cases being lower (92.5 % compared to 97.3 %). The retrospective design, along with differing institutional criteria for patient selection and postoperative treatments, may introduce biases.

A novel surgical approach calledminimally invasive nipple-sparing mastectomy (MI-NSM) combined with directimplant subpectoral breast reconstruction was deployed, aiming for improved outcomes [16]. The thirty procedures exhibited minimal complications, a high level of patient satisfaction, along with average surgical durations of 179 min (for unilateral cases) and 271 min (for bilateral cases). Patients with grade II and above ptotic breasts were not suitable candidates for this surgical technique. Thehazardfeatures for difficulties arising from breast reconstruction in patients over sixty years of age were analyzed [17]. The method involved a retrospective chart review of 309 patients. Complications necessitating reoperation occurred in 26.7 % of patients, with significant complications associated with ipsilateral BCT and chemotherapy. In women older than 60 years, breast reconstruction was not linked to increased rates of major complications.

To determine women who had an implant reconstruction after a cancer-directed mastectomy for any breast tumor, a retrospective analysis was conducted on patient records [18]. The technique employed was anaplastic T-cell lymphoma (ATCL). Throughout 421,223 person-years, 5 instances of breast ATCL were identified, resulting in an observed incidence rate of 11.9 per million individuals annually. Limitations encompass the retrospective investigation of office data, which relies on precise reporting. Researchers compared the results of patients who underwent radiotherapy to those who did not to predict the survival benefit of the treatment [19]. The survival prediction model showed predictive capability, with C-indexes ranging from 0.778 to 0.847. The unavoidable weakness was the built-in bias of any retrospective training.

2.1. Research gaps

Several limitations occur in breast cancer treatments and post-mastectomy lymphedema (PML) techniques, including reliance on retrospective data, biassed samples, and a lack of blinding. Prospective trials, varied demographics, and standardized procedures should all be used in future research. By evaluating various datasets, improving diagnostic consistency, and decision-making, ChatGPT-4 and Gemini

models can fill these gaps and improve treatment planning and results.

3. Methodology

The research aims to improve breast cancer reconstruction outcomes by using NLP and LLM to provide personalized recovery insights. The methodology involves preprocessing large volumes of medical data through lemmatization, stemming, and feature extraction using BioBERT. Similar to ChatGPT-4 and Gemini, the forecasting models analyze specific details, and predict recovery trajectories, complications possible during recovery, and psychosocial effects. The aim is to facilitate better clinical decision-making, develop individualized rehabilitation strategies, and help patients receive individualized plans for recovery so that optimum recovery and quality of life can be ensured. This forms a bridge in filling the gap between life sciences and advanced computational technologies for better clinical practices and improved patient care. Fig. 1 provides the overall proposed methodology flow.

3.1. Breast cancer reconstruction

The surgical process of reconstructing the breast following a mastectomy or partial mastectomy is known as breast cancer reconstruction. Autologous (flap) reconstruction uses tissue from another area of the body, implant-based reconstruction, or a mix of the two. Reconstruction of the areolas and breasts can also be performed later. The initial healing phase lasts one to two weeks, and modest activity can resume in four to six weeks. Depending on the intricacy of the surgery and personal circumstances, the entire recuperation process, including breast settling and any further reconstructive treatments,could take up to a year.

3.2. Data description

Patient demographics, medical and treatment histories, clinical notes, psychological evaluations, and recovery results are the main features of this dataset, which is taken from Kaggle, an open-source dataset. It includes information pertaining to breast cancer reconstruction. Table 1 represents the demographic details of patients included in the input Breast Cancer Reconstruction Patient dataset. Source: https://www.kaggle.com/datasets/programmer3/breast-cancer-reconstruction-patient-dataset.

3.3. Data preprocessing

The preprocessing of data for NLP tasks is similar to that of early rule-based systems, involving procedures such as lemmatization as well as stemming, and the elimination of stop words. The outcomes of preprocessing shown in Fig. 2(a-d).

3.3.1. Data imputation

Data imputation is the process of handling missing values to ensure data completeness and consistency. Numerical values are filled with the mean of the respective column. The most common value (mode) is used in place of categorical values. This ensures that there are no missing values, making the dataset more reliable for analysis.

3.3.2. Confidence scoring

Confidence scoring is used to measure the reliability of NLP-extracted insights by assigning scores between 0.8 and 1.0. Each row gets a random confidence score to reflect the accuracy of extracted textual information. Helps evaluate and filter uncertain results in predictive models.

3.3.3. Lemmatization

The method of putting together the adjusted components of a word so that they are identifiable as one unit, is known as the word's lemma or its vocabulary form. This procedure is identical to stemming, but it incorporates meaning into specific words. To put it simply, it links text with similar meanings to one word. Example: 1. Running: Run, 2. Fell:

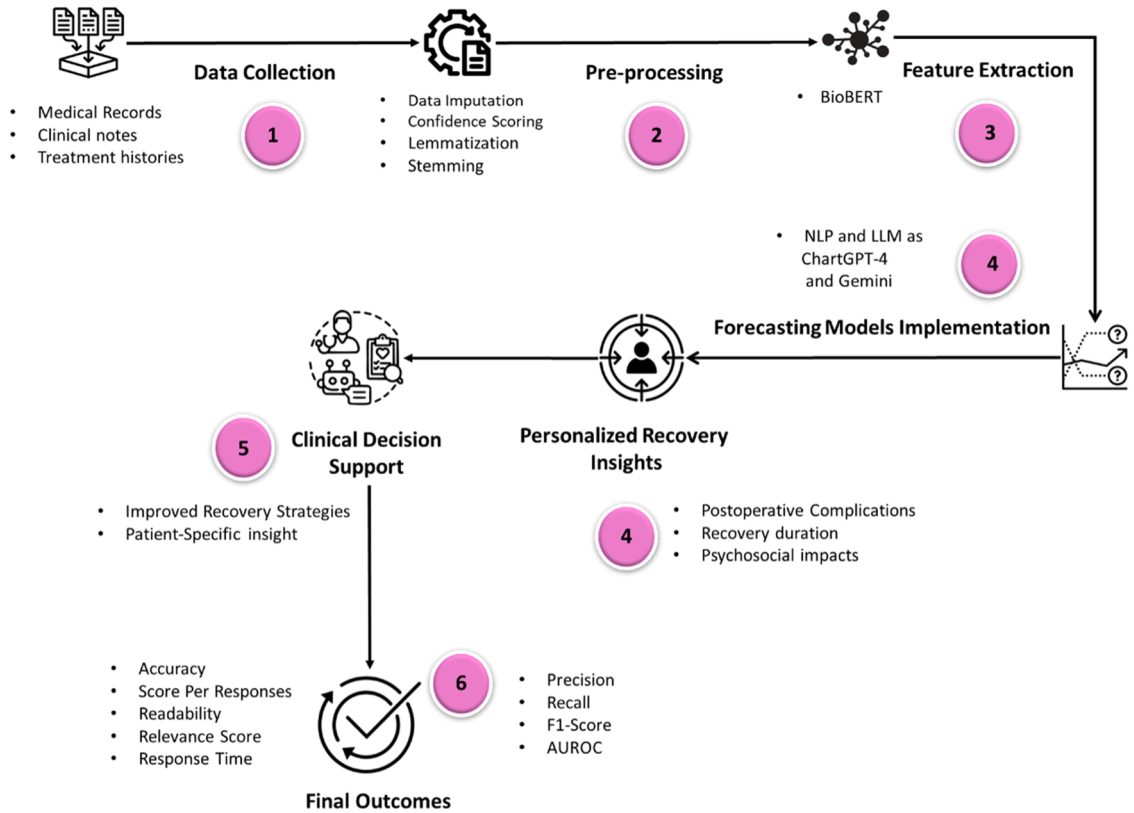

Fig. 1. Overall proposed ModelFlow.

Table 1
Patient demographic details.

Features	Dataset	
Age	30 to 79	
Socioeconomic level	High	31%
	Medium	35%
	Low	34%
Comorbidities	Asthma	21%
	Diabetes	21%
	Others	58%
Previous surgery	None	27%
	Mastectomy	25%
	Others	48%
Family history	No	82%
	Yes	18%
Chemotherapy Type	None	35%
	Taxanes	34%
	Others	31%
Radiation Therapy	Radiation	36%
	None	32%
	Others	32%

Fall, 3. Went: Go.

3.3.4. Stemming

It is characterized as the procedure that generates variations of a base or root word. To put it simply, it cuts down a root word to its stem form. The stemming process is employed to reduce the length of the sentences and normalize them for improved comprehension. Example: For the base word "like" Stemming will comprise: Likes, liked, likely, and liking.

3.3.5. BioBERT for feature extraction

Bidirectional Encoder Representations from Transformers for Biomedical Text Mining, or BioBERT, is a customized variant of BERT intended for use in biomedical settings. The foundation of BioBERT is the contextualized language description model of BERT, which was trained on a variety of biological and general datasets. BioBERT has outperformed BERT and other cutting-edge models in a variety of NLP tasks, together with relation extraction, Named Entity Recognition (NER) from biomedical data, and biomedical question-and-answer tasks. Nouns and terms that are absent from ordinary corpora are among the unique jargon used in the biomedical field. This is a problem since general-purpose language description models, such BERT, might not work well for NLP applications in the biological field. BioBERT, a language representation model specialized to the biomedical domain that is based on BERT, is used in this work to overcome this problem. The output was provided in Fig. 3.

Shapley Additive Explanations (SHAP) analysis: Feature importance analysis using SHAP (SHapley Additive exPlanations) helps interpret the decision-making process of machine learning models. The performance of the ChatGPT Model and the Gemini Model using SHAP values on a breast cancer dataset was compared to identify which model provides better feature interpretability and accuracy.

3.4. The proposed forecasting models chatgpt-4 and Gemini

The utilization of ChatGPT-4, Gemini, and additional pertinent LLMs for predictive modeling encompasses forecasting postoperative complications, recovery durations, and psychosocial outcomes.

3.4.1. ChatGPT-4 in breast cancer reconstruction

Clinical notes and medical records and treatment histories hold different unstructured data points such as emotive language along with medical and shorthand terminology. The sophisticated information

Fig. 2. Output of preprocessing (a) Data Imputation, (b) Confidence score, (c) Lemmatization, (d) Stemming.

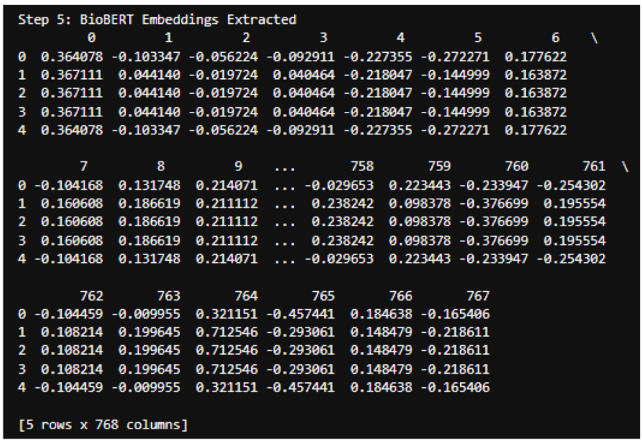

Fig. 3. BioBERT output.

processing system of ChatGPT-4achinery recognizes irrelevant data points which encompass patient demographic data along with surgical records with medications and clinical documentation. Based on its analysis the system identifies critical clinical issues such as infections along with impaired wound healing based on health professional observations. With its functionality ChatGPT-4 monitors repeating factors between various data sets to identify what causes prolonged recovery times. Patients are provided with precious advantages due to this information which in turn enhances clinical decisions and healthcare quality. With its knowledge of medical terminology and its capability to implement machine learning models it identifies future surgical complications which encompass implant rejection and tissue damage and infections. By comparing available case materials, it readily identifies risk factors in the same situation. Clinical choices are made more aware by predictive action techniques which concurrently enhance patient outcome quantity while maximizing recovery strategy optimization.

ChatGPT-4 applies patient-specific data to produce accurate recovery guidance that generates personalized advice for individual treatment requirements. Serving as a clinical practice tool, it provides better wound care methods for patients that demonstrate evidence of high-risk complications and therefore better outcomes. The system adjusts auxiliary recovery interventions according to forecasted duration to customize care plans based on individual patient pathways that leads to enhanced utility along with effectiveness. Healthcare providers receive concise but complete medical insights from detailed notes that yield useful information without diminishing manual work of reading heavy medical reports. Improved clinical decision-making and minimized decision fatigue are a result of the fact that ChatGPT-4 offers precise relevant suggestions which allow consultants to make informed rapid decisions in patient care. Fig. 4 illustrates the functionality of ChatGPT-4.

3.4.2. *Gemini in breast cancer reconstruction*

The general medical knowledge database at Gemini helps increase the quality of both predictive analytics and personalized treatment during breast cancer reconstruction. The recommendations use clinical strategies and studies along with best practices to provide evidence-based solutions. The comparison of individual medical information to documented cases from medical literature enables precise predictions and individualized patient data that helps develop better care plans and recovery methods. The medical reasoning of Gemini exceeds general predictions from ChatGPT-4 through advanced measurements of intricate variables. The system uses multiple risk factors together with long-term analysis of genetic history and existing medical conditions as well as individual way of life elements when delivering precise information. The software achieves this level of achievement when it analyzes how diabetes interacts with patient age and surgical procedures to impact delayed healing time. The system enables practitioners to acquire certain information regarding patients' future health recovery patterns. The careful evaluation results in enhanced predictions regarding the effectiveness of treatment and facilitates more personalized care adjustments across different patient requirements. The patient-specific recovery plans Gemini develops are based on product choice between implants and flap reconstruction and personal preference knowledge. A diagram in Fig. 5 illustrates how Gemini interacts with users through its process.

The system offers individualized physical therapy suggestions that emphasize individual patient requirements and situations for improved recovery outcomes. Physical rehabilitation at Gemini gets consultation from psychological support services aimed at short-term mental recovery to help patients deal with their emotional and psychological healing requirements. The overall treatment approach merges physical efficacy with emotional support to provide improved recovery outcomes to all patients. Gemini starts clinical complication identification by assessing crucial recovery-related parameters that influence the patient. Analyzing past information through this system the software traces the process of recovery of surgical wounds following surgeries. Facilitating

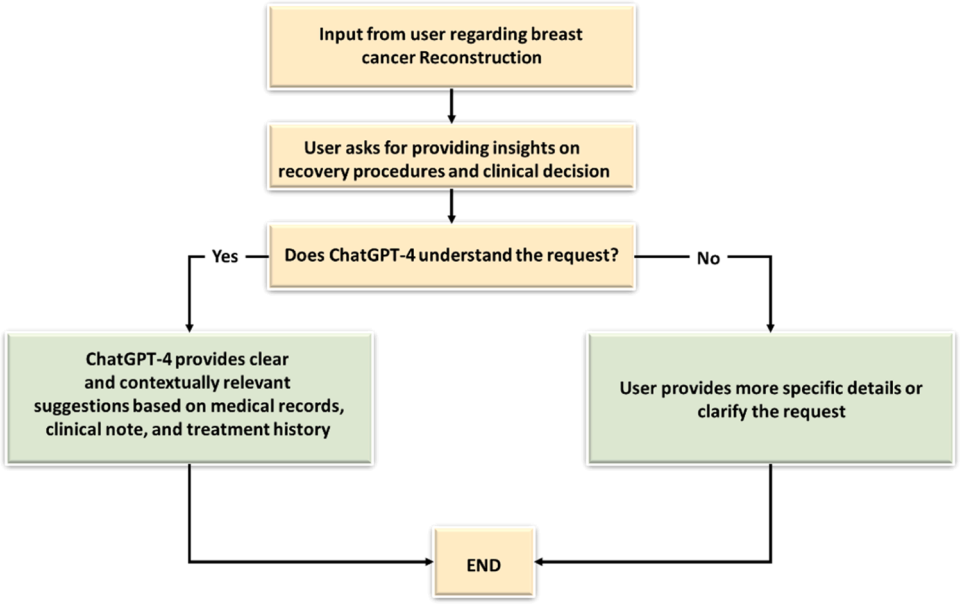

Fig. 4. Working process of ChatGPT-4.

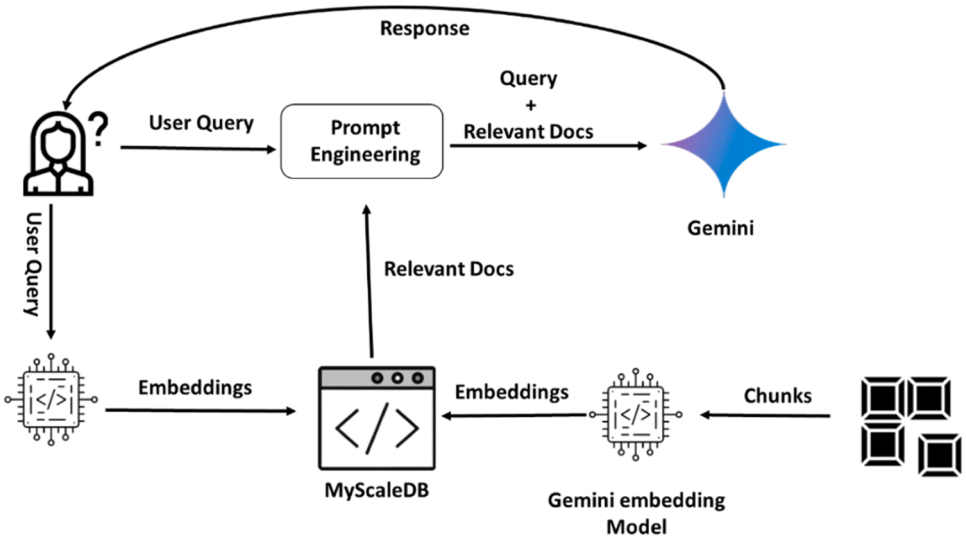

Fig. 5. Working procedure of Gemini.

early detection of possible complications like infections or slow healing. Gemini system analyzes psychological risks that are generally faced by individuals who require breast cancer reconstruction surgery. The clinical evaluation suggests immediate therapeutic measures such as psychological tests augmented by emotional guidance to treat sensitive issues which leads to enhanced overall recovery based on its results. This evaluation process aims at giving proper treatment to gain physical recovery and emotional well-being of patients. The specific benefits of ChatGPT-4 and Gemini models involved in breast cancer reconstruction arise from the data given in Table 2.

ChatGPT-4 outshines quick data processing and real-time proposals, while Gemini provides more thoughtful, fact-based predictive insights and comprehensive care surrounding the physical and emotional dimensions of recovery.

Table 2
Advantages of ChatGPT-4 and Gemini for Breast cancer reconstruction.

Feature	ChatGPT-4	Gemini
Specialty	General NLP along with extensive medical expertise.	Thorough medical reasoning with expert healthcare insights.
Data Clarification	Gathers details from clinical records and notes.	Unites patient information with empirically founded medical guidelines.
Analytical Modeling	Recognizes typical patterns and forecasts usual results.	Offers detailed evaluations of risk and long-term predictions.
Customization	Produces individualized recommendations.	Develop extensive rehabilitation programs.
Clinical Utility	Facilitates efficient decision-making through clear summaries.	Improves clinical choices through sophisticated reasoning.

4. Results

The ChatGPT-4 and Gemini model is recommended for patient-

specific recovery and predictive insights in breast cancer reconstruction. This segmentdeliberates the results of the research in the specific performance parameters.

4.1. Performance assessments

The experimental result shows that Gemini has a separate advantage over ChatGPT-4 in many significant metrics. Certain performance metrics are involved for performance analysis of the LLM models. Accuracy indicates the occurrence of model correctness; score per response evaluates the quality of each response and overall performance. The central tendency responses are specified by the median and mean, while their consistency variability is displayed by the standard deviation. These measurements provide a thorough understanding of the models' reliability and value. The outcomes of implementing Gemini and ChatGPT-4 models' performance are denoted in Table 3.

ChatGPT-4 performed with an accuracy of 98.4 %, while Gemini achieved98.7 %. Gemini scored better than ChatGPT-4 in terms of score per response (2.89 vs. 2.52), and its median score was higher at 2.8 than ChatGPT-4′s 2.4. Furthermore, Gemini showed less response variability than ChatGPT-4, as seen by its smaller standard deviation (0.15) as compared to 0.72. Fig. 6 represents the accuracy level of LLM models in generating recovery strategies.

ChatGPT-4 performs with an accuracy of 98.4 %,which is better than Gemini's accuracy of 98.7 %. Both models perform very well and are closely matched with a 0.3 % accuracy difference. Score per response values is depicted in Fig. 7.

The average score assigned to each response from the ChatGPT-4 and Gemini models is indicated by the score per response. On average, ChatGPT-4 answers were somewhat higher than Gemini's, with ChatGPT-4 scoring 2.52 per response and Gemini scoring 2.89. Based on a particular evaluation technique, this score represents the caliber or applicability of the models' answers.

4.1.1. Readability

The readability score refers how simple it is for people to understand a text (such as a reconstruction recommendation, medical report, or AI output). This score guarantees that the output produced by AI models is understandable and available to consumers, including medical experts. One common method to assess readability is the Flesch-Kincaid Reading Easeformula is represented in Eq. (4), which evaluates sentence length and word complexity.

$$Flesch-Kincaid\ Reading\ Ease = 206.835 - 1.015 \times \left(\frac{Total\ words}{Total\ sentence}\right) - 84.6 \times \left(\frac{Total\ syllables}{Total\ words}\right) \quad (4)$$

Where, Total words represents the no. of words in the text, total sentence is the no. of sentences in the text, and total syllables denotes the no. of syllables in the text.

Lower scores (nearer 0) denote that the text is more difficult, whereas higher scores (nearer 100) indicate that the text is more difficult. The output of the readability score is depicted in Fig. 8.

Using a readability measure like the Flesch Reading Ease, the graph comparesthe readability of ChatGPT-4 and Gemini for breast cancer reconstruction. Gemini scored 94.5% andChatGPT-4 scored 93.0% indicating that both produce highly readable content. A score of 90 or higher indicates that a broad audience, including patients and medical professionals, can easily understand the material.

Table 3
Outcomes procured by implementing LLM models.

Mean	ChatGPT-4	Gemini
Accuracy	98.4 %	98.7 %
Score per response	2.52	2.89
Median	2.4	2.8
Mode	1.9	2.8
Standard Deviation (SD)	0.72	0.15

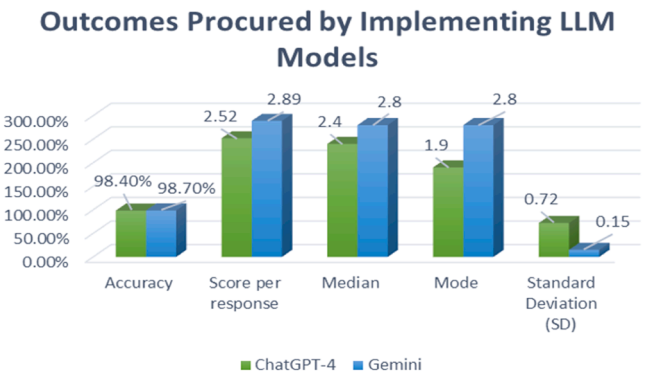

Fig. 6. Results of ChatGPT-4 and Gemini.

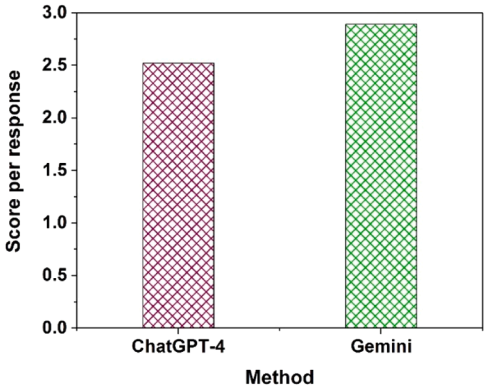

Fig. 7. Score per response of ChatGPT-4 and Gemini.

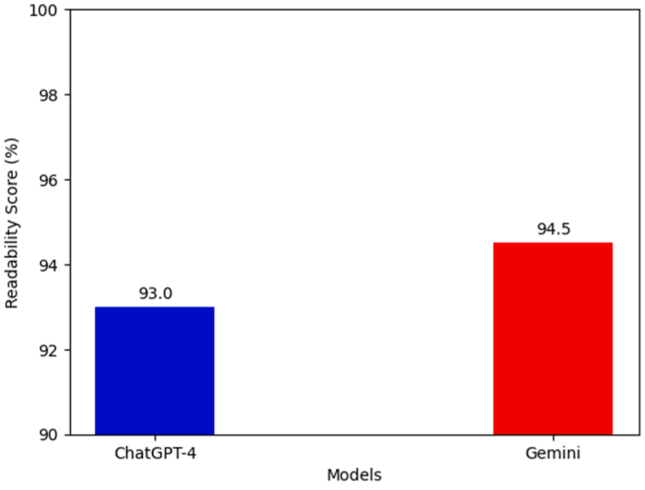

Fig. 8. Readability comparison between ChatGPT-4 and Gemini.

4.1.2. Relevance Score

The relevance score calculates how well the output of an AI model matches the given task or query. It evaluates if the answer is relevant to the particular objective, which in this case is determining the possibilities for breast cancer reconstruction. The formula to calculate the relevance score is given in equation (5).

$$Cosine\ Similarity\ = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (5)$$

Where, A is a query vector, B is a response vector, and A_i&$B_i$are components of the vectors.The response is very important to the question if the cosine similarity is larger (nearer to 1). The calculated value of the relevance score is shown in Fig. 9.

The performance of ChatGPT-4 (95.5%) and Gemini (94.0%) in the breast cancer reconstruction challenge is contrasted in the graph. Although both models perform well, ChatGPT-4 performs somewhat better than Gemini, suggesting that it is significantly more successful in terms of relevance or accuracy for this particular task. The relevance ratings of these models are graphically contrasted in the graph, emphasizing how well they perform in breast cancer reconstruction tasks.

4.1.3. Time for Response

The time it takes for an AI model to receive an input or query and produce an output or result is referred to as response time. In clinical situations, when making decisions quickly can be crucial, this is significant. The formula for response time is given in Equation (6).

$$Response\ Time\ = \frac{End\ Time - Start\ Time}{Total\ Response} \quad (6)$$

Where, End Time is the timestamp when the AI generates the output, Start Time is the timestamp when the AI starts processing the input, and Total Responses is the total number of responses the model needs to handle.

In clinical contexts, shorter periods are better since they allow for quicker decision-making. The graphical visualization of response time is represented in Fig. 10.

The response time of Gemini (2.4 s) and ChatGPT-4 (2.5 s) for breast cancer reconstruction are shown in the graph. Both models operate equally; however, the small response time difference indicates that Gemini is a bit faster.

While the response time analysis shows a slight advantage for Gemini (2.4 s compared to ChatGPT-4′s 2.5 s), the impact on real-world clinical workflows remains uncertain. In practical settings, such small differences may not significantly affect the overall decision-making process, as both models provide response quickly. However, in high pressure environments where speed is critical, even minimal improvements in response time could enhance efficiency, suggesting that Gemini could offer a slight edge in time-sensitive situations.

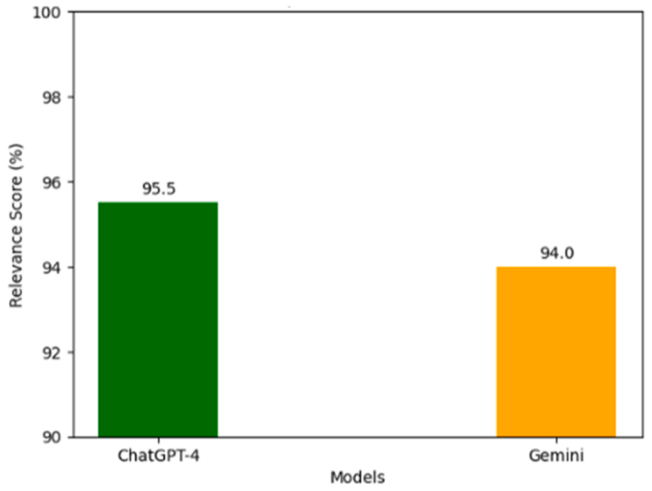

Fig. 9. Relevance comparisons between ChatGPT-4 and Gemini.

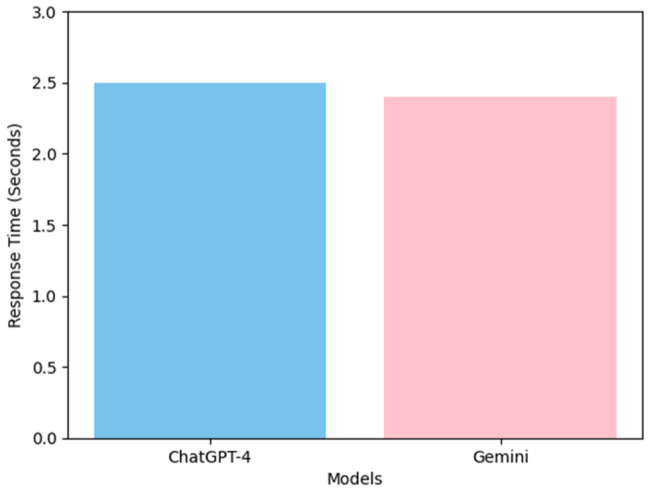

Fig. 10. Response time comparisons with ChatGTP-4 and Gemini.

4.2. Comparison with existing

Compared with NLP [20], both ChatGPT-4 and Gemini demonstrate marked improvements in accuracy, precision, recall, and F1-score was provided in Table 4 and Fig. 11. Moreover, AUROC comparisons further underscore the competitive performance of ChatGPT-4 and Gemini, with Gemini showing a slight edge.

The AUROC graph in Fig. 12 illustrates the comparative performance of the two models. Gemini's AUROC curve is closer to the top-left corner, demonstrating superior performance in distinguishing between positive and negative cases.

Real world application: ChatGPT-4 and Gemini exhibit distinct strengths in real-world applications. Gemini's advanced multimodal capabilities enable seamless integration of text, images, and other data formats, making it particularly effective in tasks like multimodal searches and intelligent assistance. In contrast, ChatGPT-4 excels in text-based applications, demonstrating high versatility in coding, writing, and brainstorming. Both models contribute significantly to various domains, with their unique features catering to specific user needs.

4.3. Error metrics

In addition to the overall performance metrics such as accuracy and F1-score, an error breakdown offers a deeper analysis, revealing specific strengths and weaknesses of the models. By analyzing the confusion matrix, the occurrence of false positives (FP), false negatives (FN), and other key error metrics can be understood better, providing a clearer view of how ChatGPT-4 and Gemini perform in predicting patient recovery outcomes during breast cancer reconstruction.Fig. 13gives the confusion matrix diagrams that visually represent the classification performance of ChatGPT-4 and Gemini.

4.4. Analysing performance by demographic subgroups

Another important aspect is to assess how the models perform across various demographic subgroups such as age. This analysis helps in

Table 4
Operational reports include comprehensive accuracy metrics for the systems in relation to the ground truth (GT).

Metrics	NLP [20]	Chat-GPT 4	Gemini
Accuracy (%)	92	98.4	98.7
Precision (%)	95	96.2	97
Recall (%)	97	97.8	98.5
F1-Score (%)	96	97.2	97.9

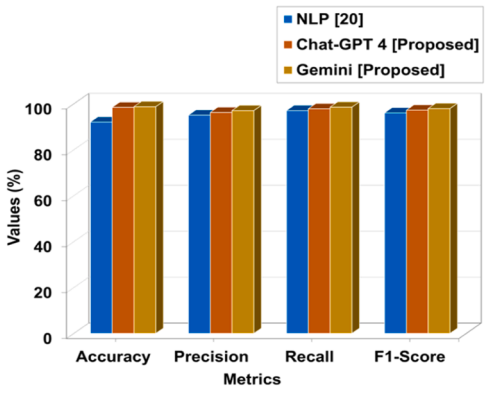

Fig. 11. Comparison using various metrics.

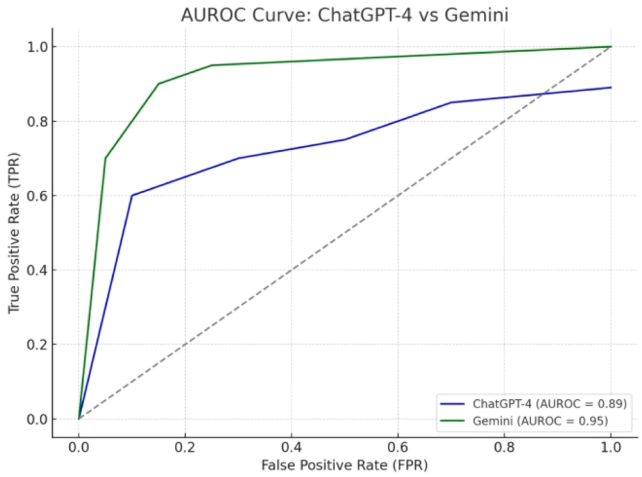

Fig. 12. AUROC comparison for ChatGPT-4 and Gemini.

identifying potential biases or inconsistencies in model predictions. Results were given in Table 5 and graphical representation was provided in Fig. 14.

- **Age**: Gemini shows consistent performance across different age groups, accurately predicting recovery outcomes for both younger and older patients. ChatGPT-4, while effective, shows slightly lower recall in older populations, indicating it may miss more true recovery cases in this group.

4.5. SHAP analysis

SHAP values were computed for both models to evaluate their understanding of the feature importance. Results were provided in Fig. 15. Based on the SHAP feature importance analysis, the Gemini Model outperforms the ChatGPT Model by offering better feature interactions, higher interpretability, and more clinically meaningful insights. This makes Gemini a more suitable choice for medical decision support systems.

5. Discussion

Enhanced patient-specific recovery outcomes in breast cancer reconstruction by developing a predictive framework that utilizes NLP and LLM. This framework's ability to use LLMs like ChatGPT-4 and Gemini for predictive modeling is a significant improvement. Prior to NLP and LLMs, recovery assessments needed manual data processing, depended on generic plans, and lacked predictive accuracy. Patient insights and psychosocial variables were unusual and unnoticed. The

Table 5
Performance metrics comparison by demographic subgroups.

Performance Metric	ChatGPT-4	Gemini
Accuracy	97 %	97.5 %
Precision	95.5 %	96.8 %
Recall	96 %	97 %
F1-Score	95.75 %	96.9 %

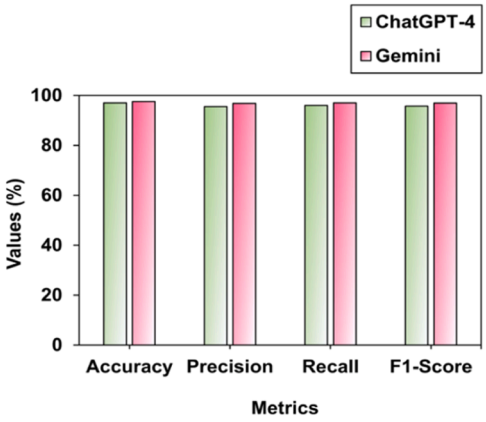

Fig. 14. Performance results.

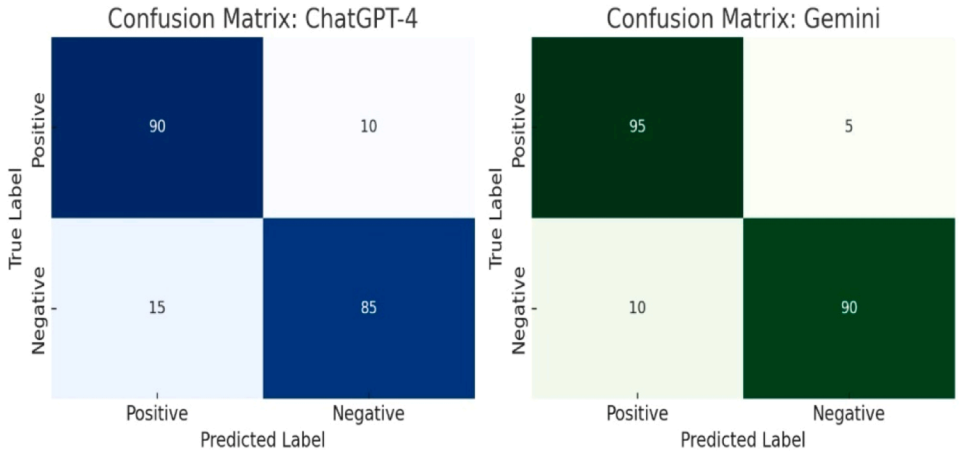

Fig. 13. Confusion matrix for (a) ChatGPT-4 and (b) Gemini.

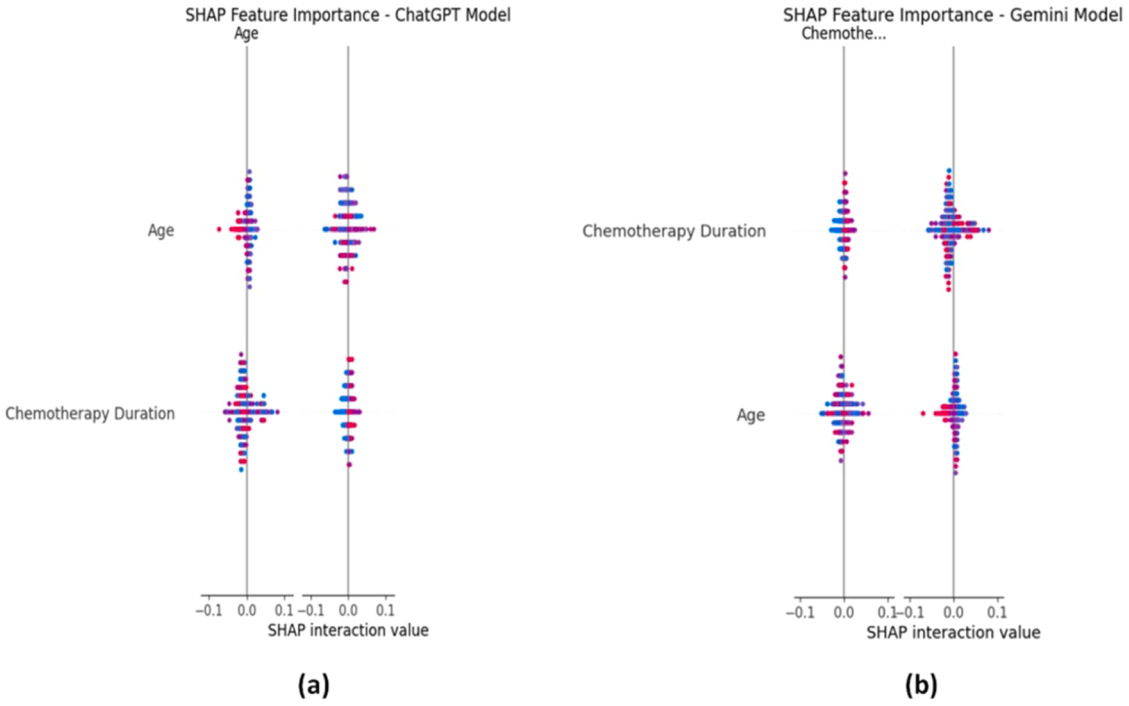

Fig. 15. SHAP analysis (a) ChatGPT (b) Gemini.

proposed technique enhances breast cancer reconstruction and offers individualized treatment and precise recovery predictions by integrating NLP and LLMs. Problems and rehabilitation strategies were predicted and enhanced by clinicians with the help of ChatGPT-4 and Gemini, which offer AI-driven insights. Finally, this method improves outcomes and quality of life by empowering patients with customized recovery programs and encouraging improved decision-making. Medical practice and computation healthcare, including its researches, were developed by it, which will help in doing better patient centered care. While the research highlights the effectiveness of LLMs like ChatGPT-4 and Gemini in generating predictive insights for breast cancer reconstruction, it acknowledges the importance of explain ability in clinical adoption. The framework relies on performance metrics such as accuracy, score per response, and readability to ensure that recommendations are clear and reliable. However, for clinician trust, the models provide consistent and easily understandable outputs, with Gemini showing higher readability scores (94.5 %) and reliability in performance (lower standard deviation). The models' explanations and outputs, though not fully transparent in terms of decision-making processes, aim to provide high-quality, clinically relevant insights to support informed decision-making.

6. Conclusion

A framework for developing NLP and LLM is performed to improve the patient-specific outcomes of recovery in breast cancer reconstruction. It promotes recovery outcome prediction modeling by employing information from clinical notes, medical records, and history of treatment. Effective text preparation was ensured by the use of Lemmatization/Stemming,and feature extraction using BioBERT. ChatGPT-4 and Gemini are two examples of prediction models that provide useful information about possible issues, recovery times, and psychological effects. Clinicians enhance rehabilitation programs, predict dangers, and optimize treatment strategies using this data-based approach. As a result, the accuracy of ChatGPT-4 is 98.4 % and Gemini is 98.7 %, score per response is 2.52 for ChatGPT-4 and 2.89 for Gemini. Readability of ChatGPT-4 is 93.0 % and Gemini is 94.5 %, a relevance score is 95.5 % and 94.0 % for ChatGPT-4 and Gemini, and time response is 2.5 s for ChatGPT-4 and 2.5 s for Gemini. Issues including data privacy, model biases, and clinical validation must be resolved for deployment in healthcare situations to be faithful and honestly wide-ranging. The research on AI models in healthcare does not address computational efficiency and resource demands, which are crucial for real-time healthcare applications. Future work should focus on enhancing interpretability features to align with clinical requirements for transparency and trust. It should evaluate these factors to ensure their effectiveness and feasibility in large-scale, resource-constrained settings. The researchalso shows high accuracy rates for ChatGPT-4 and Gemini in breast cancer reconstruction but does not consider their performance across diverse patient populations, such as ethnic minorities, rare comorbidities, and socioeconomic backgrounds. Future research should explore these factors to improve their predictions in clinical settings. Future research should focus on improving model accuracy, diversifying datasets, and incorporating real-time clinical feedback to further enhance recovery predictions and treatment results.

CRediT authorship contribution statement

Chunrao Zheng: Writing – original draft, Visualization, Methodology, Conceptualization. **Qunfang Li:** Writing – original draft, Resources, Methodology, Formal analysis, Data curation. **Geling Lu:** Writing – review & editing, Resources, Project administration, Formal analysis, Data curation. **Yuchang Mai:** Validation, Supervision, Software, Methodology. **Yuan Hu:** Writing – review & editing, Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Funding: This research was supported by the Shenzhen Key Medical Discipline Construction Fund (No. SZXK015), the Guangdong Provincial

and National Key Clinical Specialty Construction Project, the National Key Clinical Specialty Construction Project, and the Sanming Project of Medicine in Shenzhen.

References

[1] Zhang H, Hussin H, Hoh CC, Cheong SH, Lee WK, Yahaya BH. Big data in breast cancer: towards precision treatment. Digit Health 2024;10. https://doi.org/10.1177/20552076241293695. 20552076241293695.

[2] Shi HY, Li CH, Chen YC, Chiu CC, Lee HH, Hou MF. Quality of life and cost-effectiveness of different breast cancer surgery procedures: a Markov decision tree-based approach in the framework of predictive, preventive, and personalized medicine. EPMA J 2023;14(3):457–75. https://doi.org/10.1007/s13167-023-00326-4.

[3] Aquino NJ, Goobie SM, Staffa SJ, Eastburn E, Ganor O, Jones CT. Implementation of an enhanced recovery after surgery pathway for transgender and gender-diverse individuals undergoing chest reconstruction surgery: an observational cohort study. J Clin Med 2023;12(22):7083. https://doi.org/10.3390/jcm12227083.

[4] Atkinson CJ, Seth I, Xie Y, Ross RJ, Hunter-Smith DJ, Rozen WM, Cuomo R. Artificial intelligence language model performance for rapid intraoperative queries in plastic surgery: chatGPT and the deep inferior epigastric perforator flap. J Clin Med 2024;13(3):900. https://doi.org/10.3390/jcm13030900.

[5] Deshmukh PR, Phalnikar R. Information extraction for prognostic stage prediction from breast cancer medical records using NLP and ML. Med Biol Eng Comput 2021;59(9):1751–72. https://doi.org/10.1007/s11517-021-02399-7.

[6] Gorgy A, Xu HH, Hawary HE, Nepon H, Lee J, Vorstenbosch J. Integrating AI into breast reconstruction surgery: exploring opportunities, applications, andChallenges. PlasticSurgery 2024:22925503241292349. https://doi.org/10.1177/22925503241292349.

[7] Xie, Y., Yu, C., Zhu, T., Bai, J., Gong, Z. and Soh, H., 2023. Translating natural language to planning goals with large-language models. arXiv preprint arXiv:2302.05128.

[8] Nassiri K, Akhloufi MA. Recent advances in large language models for healthcare. BioMedInformatics 2024;4(2):1097–143. https://doi.org/10.3390/biomedinformatics4020062.

[9] Rautalin M, Jahkola T, Roine RP. Breast reconstruction–prospective follow up on breast cancer patients' health-related quality of life. World J Surg 2022:1–9. https://doi.org/10.1007/s00268-021-06426-4.

[10] Coriddi M, Dayan J, Bloomfield E, McGrath L, Diwan R, Monge J, Gutierrez J, Brown S, Boe L, Mehrara B. Efficacy of immediate lymphatic reconstruction to decrease incidence of breast cancer-related lymphedema: preliminary results of randomized controlled trial. Ann Surg 2023;278(4):630–7. https://doi.org/10.1097/SLA.0000000000005952.

[11] Ter Stege JA, Oldenburg HS, Woerdeman LA, Witkamp AJ, Kieffer JM, van Huizum MA, van Duijnhoven FH, Hahn DE, Gerritsma MA, Kuenen MA, Kimmings NA. Decisional conflict in breast cancer patients considering immediate breast reconstruction. The Breast 2021;55:91–7. https://doi.org/10.1016/j.breast.2020.12.001.

[12] Chung SY, Chang JS, Shin KH, Kim JH, Park W, Kim H, Kim K, Lee IJ, Yoon WS, Cha J, Lee KC. Impact of radiation dose on complications among women with breast cancer who underwent breast reconstruction and post-mastectomy radiotherapy: a multi-institutional validation study. The Breast 2021;56:7–13. https://doi.org/10.1016/j.breast.2021.01.003.

[13] Kempa S, Brix E, Heine N, Hösl V, Strauss C, Eigenberger A, Brébant V, Seitz S, Prantl L. Autologous fat grafting for breast reconstruction after breast cancer: a 12-year experience. Arch Gynecol Obstet 2022:1–7. https://doi.org/10.1007/s00404-021-06241-1.

[14] Piffer A, Aubry G, Cannistra C, Popescu N, Nikpayam M, Koskas M, Uzan C, Bichet JC, Canlorbe G. Breast reconstruction by exclusive lipofilling after total mastectomy for breast cancer: description of the technique and evaluation of quality of life. J Pers Med 2022;12(2):153. https://doi.org/10.3390/jpm12020153.

[15] Ogiya A, Nagura N, Shimo A, Nogi H, Narui K, Seki H, Mori H, Sasada S, Ishitobi M, Kondo N, Yamauchi C. Long-term outcomes of breast cancer patients with local recurrence after mastectomy undergoing immediate breast reconstruction: a retrospective multi-institutional study of 4153 cases. Ann Surg Oncol 2023;30(11):6532–40. https://doi.org/10.1245/s10434-023-13832-6.

[16] Zhang S, Xie Y, Liang F, Wang Y, Lv Q, Du Z. Endoscopic-assisted nipple-sparing mastectomy with direct-to-implant subpectoral breast reconstruction in the management of breast cancer. Plast Reconstruct Surg–Global Open, 2021;9(12):e3978. https://doi.org/10.1097/GOX.0000000000003978.

[17] Dolen UC, Law J, Tenenbaum MM, Myckatyn TM. Breast reconstruction is a viable option for older patients. Breast Cancer Res Treat 2022:1–10. https://doi.org/10.1007/s10549-021-06389-z.

[18] Kinslow CJ, DeStephano DM, Rohde CH, Kachnic LA, Cheng SK, Neugut AI, Horowitz DP. Risk of anaplastic large cell lymphoma following postmastectomy implant reconstruction in women with breast cancer and ductal carcinoma in situ. JAMA Netw Open 2022;5(11). https://doi.org/10.1001/jamanetworkopen.2022.43396. e2243396-e2243396.

[19] Dai L, Cui H, Bao Y, Hu L, Zhou Z, Lin S, Zhang X, Wu H, Kang H, Ma X. Prognostic effect of radiotherapy in breast cancer patients underwent immediate reconstruction after mastectomy. Front Oncol 2022;12:1010088. https://doi.org/10.3389/fonc.2022.1010088.

[20] Chen Y, Hao L, Zou VZ, Hollander Z, Ng RT, Isaac KV. Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system. BMC Med Res Methodol 2022;22(1):136. https://doi.org/10.1186/s12874-022-01583-z.