

Research Article

Analysing Breast Cancer Risk Through Food Habits Using CNN and LSTM Models

V Thangavel^{1*} and Venkatesan²

¹Francis Institute of Management and Research, Mumbai, India.

*Corresponding Author: V Thangavel, Francis Institute of Management and Research, Mumbai, India.

²Department of Computer Science, RV Government Arts College, Chengalpattu, India.

Received: 📅 2026 Feb 16

Accepted: 📅 2026 Mar 10

Published: 📅 2026 Mar 19

Abstract

Breast cancer is a major global health concern and one of the leading causes of death among women. Its risk is influenced by both genetic and environmental factors, with diet recognized as a key modifiable contributor. This study focuses on the role of vegetarian and non-vegetarian food habits in influencing breast cancer progression and applies deep learning models to predict risk outcomes.

Objective: The primary objective is to analyze how different dietary patterns, particularly plant-based versus animal-based diets, affect breast cancer risk and progression.

Methods: A dataset of breast cancer patients was classified into vegetarian and non-vegetarian groups. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models were employed to predict patient risk levels and identify hidden patterns linking dietary intake with cancer growth.

Findings: The results reveal that non-vegetarian diets, especially those high in fats, red meat, and processed foods, are associated with increased cell growth and higher cancer risk. In contrast, vegetarian diets rich in fiber, antioxidants, and phytochemicals demonstrate a protective effect. CNN achieved the highest accuracy, while LSTM effectively captured sequential dietary behaviors. These findings confirm that deep learning can serve as a valuable tool for diet-based risk prediction in breast cancer patients.

Keywords: Breast Cancer, Dietary Habits, Vegetarian Diet, Deep Learning, Cancer Risk Prediction

1. Introduction

Breast cancer is a widespread and serious health condition that affects millions of women worldwide, representing one of the leading causes of cancer-related mortality. Studies indicate that approximately one in eight women will develop breast cancer at some point in their lives, highlighting the urgency of understanding both its causes and preventive strategies [1]. The development of breast cancer is influenced by a combination of genetic factors, such as mutations in the BRCA1 and BRCA2 genes, as well as environmental and lifestyle factors. Among lifestyle influences, dietary habits play a significant role in either accelerating or slowing down the progression of the disease. Nutrition affects key biological processes, including inflammation, oxidative stress, hormone regulation, and cellular metabolism, all of which can impact tumor growth and patient outcomes.

Research has demonstrated that the type of diet consumed may influence breast cancer risk and progression. Non-vegetarian diets, especially those high in red meat, processed foods, and saturated fats, have been linked to increased inflammation, oxidative DNA damage, and abnormal cell proliferation [2]. Cooking methods such as grilling and frying at high temperatures can produce harmful compounds, including heterocyclic amines and polycyclic aromatic hydrocarbons, which are known to have carcinogenic effects. In contrast, vegetarian diets, rich in fruits, vegetables, legumes, and whole grains, provide essential nutrients such as vitamins, minerals, fibre, and bioactive compounds like polyphenols and flavonoids. These nutrients support the body's defence mechanisms by reducing oxidative damage, regulating estrogen metabolism, promoting apoptosis in abnormal cells, and enhancing immune responses [3]. Furthermore, fibre-rich plant-based diets help maintain a healthy gut microbiome, which has

been linked to lower systemic inflammation and improved regulation of hormones, including estrogen, a critical factor in many breast cancers.

Beyond general health benefits, specific biological mechanisms explain why diet influences cancer development. Diets high in saturated fats and processed meats can activate inflammatory pathways, increase estrogen levels, and trigger oxidative stress, all of which promote tumor growth and metastasis. On the other hand, plant-based diets provide compounds that can suppress inflammatory cytokines, regulate gene expression, and support DNA repair. Observational studies have reported lower breast cancer incidence and improved survival rates in populations consuming predominantly vegetarian diets compared to those with high meat consumption. These findings emphasise that dietary choices are not only supportive measures but also integral components in reducing cancer risk and managing disease progression.

In addition to the nutritional perspective, advances in artificial intelligence, particularly deep learning, offer new opportunities to predict disease outcomes and identify high-risk individuals. Techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are capable of processing large, complex datasets to uncover patterns that may be missed by conventional analysis methods (LeCun, Bengio, and Hinton). CNNs are effective in detecting spatial patterns and relationships within data, while LSTMs are designed to handle sequential information, making them particularly suitable for analysing dietary habits and longitudinal patient records. By integrating deep learning models with patient nutrition and clinical data, researchers can develop predictive frameworks that not only classify risk levels but also inform personalised dietary interventions aimed at minimising breast cancer progression.

This study aims to investigate the relationship between vegetarian and non-vegetarian diets and breast cancer progression using patient dietary datasets. By applying CNN and LSTM models, the study seeks to predict risk levels, detect correlations between diet and tumor growth, and provide insights into how nutrition influences disease outcomes. The findings from this research are expected to guide personalised dietary recommendations and contribute to the development of AI-assisted tools for clinical decision-making in oncology. This integration of nutritional science with machine learning has the potential to enhance preventive strategies and improve overall patient management, emphasising the critical role of lifestyle factors in conjunction with advanced computational approaches. In summary, understanding the impact of dietary habits on breast cancer risk is essential for effective disease prevention and management.

While diets high in red meat and processed foods are associated with increased risk and aggressive tumor behaviour, plant-based diets rich in fibre, antioxidants, and

micronutrients can provide protective effects. The use of deep learning models provides a powerful approach to analysing complex dietary and clinical datasets, enabling accurate risk prediction and the development of targeted nutritional interventions. This study aims to bridge the gap between dietary research and predictive analytics, demonstrating how personalised nutrition strategies, supported by AI, can contribute to better breast cancer outcomes.

2. Literature Review

Breast cancer continues to be a major area of medical research due to its prevalence, complex aetiology, and high mortality rate among women worldwide. Epidemiological studies have consistently demonstrated that breast cancer is influenced by both genetic predispositions and environmental factors. Mutations in genes such as BRCA1 and BRCA2 significantly increase the lifetime risk of developing breast cancer, while hormonal factors, reproductive history, and exposure to environmental toxins contribute to disease incidence [1]. However, lifestyle-related factors, including diet, physical activity, and obesity, have received increasing attention as modifiable determinants that can either enhance or reduce breast cancer risk. In particular, diet has emerged as a critical area of study because it directly affects metabolic processes, inflammation, oxidative stress, and hormone regulation, all of which play important roles in tumor initiation and progression.

Several studies have examined the relationship between dietary habits and breast cancer outcomes. Diets rich in red and processed meats, saturated fats, and refined carbohydrates have been associated with a higher risk of breast cancer and more aggressive tumor growth [2]. Compounds generated during high-temperature cooking of meat, such as heterocyclic amines and polycyclic aromatic hydrocarbons, have been identified as potential carcinogens. These compounds can induce DNA damage, promote oxidative stress, and trigger abnormal cell proliferation. Additionally, high-fat diets are linked to elevated estrogen levels, which are particularly relevant in estrogen receptor-positive breast cancer. Conversely, plant-based diets, including vegetarian and vegan diets, are rich in fiber, vitamins, minerals, and phytochemicals such as polyphenols and flavonoids, which provide protective effects against tumor development. Fiber consumption has been shown to improve gut microbiota composition, reduce systemic inflammation, and aid in estrogen regulation, thereby lowering cancer risk [3]. The inclusion of micronutrients such as vitamin C, vitamin E, selenium, and zinc in plant-based diets supports DNA repair mechanisms and inhibits tumor angiogenesis, further demonstrating the potential protective role of vegetarian diets.

Beyond epidemiological evidence, research has focused on the molecular mechanisms through which dietary components influence breast cancer risk. Phytochemicals in fruits, vegetables, and legumes have been shown to modulate gene expression, reduce oxidative stress, and induce apoptosis in abnormal cells. Certain bioactive compounds

can inhibit the proliferation of cancer cells by regulating signalling pathways involved in inflammation and cell cycle control. On the other hand, high intake of saturated fats and processed meats is associated with activation of pro-inflammatory pathways, increased oxidative DNA damage, and stimulation of angiogenesis, which collectively contribute to tumor progression. Studies suggest that populations with predominantly plant-based diets exhibit lower incidence rates of breast cancer and improved prognosis, indicating that diet can serve as a key preventive and therapeutic measure. Furthermore, several meta-analyses have confirmed the protective effects of vegetarian diets, highlighting reduced risk of hormone receptor-positive breast cancers in women adhering to high-plant diets compared to those consuming predominantly animal-based foods.

The integration of computational techniques and artificial intelligence in breast cancer research has further expanded the scope of dietary studies. Deep learning algorithms, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have been widely applied in healthcare to predict disease risk, classify tumor types, and analyse complex patient data (LeCun, Bengio, and Hinton). CNNs are particularly effective in identifying spatial and structural patterns, making them suitable for imaging-based diagnostics and pattern recognition within high-dimensional datasets. LSTM networks, designed to handle sequential and time-series data, are well-suited for analysing longitudinal dietary records, patient health histories, and temporal patterns in tumor progression. Combining dietary datasets with deep learning models allows researchers to detect hidden correlations between nutrition and cancer risk that are difficult to uncover through traditional statistical approaches. Recent studies employing these techniques have demonstrated high predictive accuracy in identifying high-risk patients based on food intake, body mass index, hormone levels, and lifestyle factors, emphasising the potential of AI-assisted models in personalised nutrition and clinical decision-making.

Several studies have specifically focused on the predictive power of deep learning models in breast cancer risk assessment using dietary and lifestyle data. For example, CNN-based models can classify patients into risk categories by analysing complex interactions between multiple dietary components, while LSTM networks capture sequential consumption patterns to understand long-term dietary impacts. These approaches provide actionable insights, enabling healthcare professionals to design targeted interventions and recommend personalised dietary modifications to reduce cancer risk. Furthermore, the use of AI in dietary assessment helps overcome limitations of self-reported data by identifying patterns that may not be evident through conventional analysis. Integrating nutritional data with predictive modelling represents a transformative approach, linking lifestyle factors with technology-driven solutions to improve patient outcomes.

In conclusion, existing literature emphasises that both

dietary habits and computational modelling play critical roles in understanding and managing breast cancer risk. Non-vegetarian diets, high in fats, red meats, and processed foods, contribute to oxidative stress, inflammation, and accelerated tumor growth, while vegetarian diets provide protective nutrients that regulate hormones, enhance immunity, and inhibit cancer progression. Deep learning models, including CNN and LSTM, offer powerful tools to analyse complex dietary and clinical datasets, providing accurate risk prediction and supporting personalised nutrition recommendations. The convergence of nutritional science and artificial intelligence creates a promising framework for precision oncology, enabling interventions that reduce breast cancer risk and improve long-term patient prognosis. This body of research underscores the importance of integrating lifestyle modification with advanced computational techniques to develop effective, evidence-based strategies for cancer prevention and management.

3. Methodology

This study employs a quantitative research design that integrates dietary analysis with predictive modelling using deep learning algorithms to examine the impact of vegetarian and non-vegetarian diets on breast cancer risk and progression. The research methodology is divided into several stages, including data collection, preprocessing, model selection, training, evaluation, and result interpretation.

3.1. Data Collection

The dataset used in this study comprises dietary and clinical information from breast cancer patients, collected from hospital records, nutrition surveys, and publicly available healthcare datasets. Patients were classified into two categories based on their dietary habits: vegetarian and non-vegetarian. Additional patient attributes such as age, body mass index (BMI), hormonal receptor status, tumor stage, and treatment history were also included to provide a comprehensive understanding of the factors influencing breast cancer progression. The inclusion criteria required that participants have complete dietary records and confirmed breast cancer diagnoses.

3.2. Data Preprocessing

Before applying deep learning models, the collected data underwent preprocessing to ensure accuracy and consistency. Missing values were handled using appropriate imputation techniques, while categorical variables, such as diet type and tumor stage, were encoded using one-hot encoding. Numerical features, including age and BMI, were normalised to a common scale to improve model convergence. Outliers were identified and treated using statistical methods to prevent skewed predictions. Feature selection was performed to identify the most relevant variables influencing breast cancer risk, with particular focus on dietary patterns and nutrient intake levels.

3.3. Model Selection and Implementation

Two deep learning algorithms were implemented to analyse the dataset: Convolutional Neural Networks (CNN) and

Long Short-Term Memory (LSTM) networks. CNNs were selected due to their ability to identify complex patterns and interactions among features, while LSTM networks were chosen for their capability to capture sequential and temporal relationships in dietary data. Both models were constructed using Python and TensorFlow frameworks, with architecture optimisation achieved through hyperparameter tuning, including the number of layers, neurons, learning rate, and activation functions. Dropout layers were incorporated to reduce overfitting, and the models were trained using a supervised learning approach with labelled data indicating risk categories and tumor progression outcomes.

3.4. Model Training and Evaluation

The dataset was divided into training, validation, and testing subsets using a 70:15:15 split to ensure reliable model evaluation. During training, the models were iteratively updated using backpropagation and gradient descent optimisation techniques. Model performance was evaluated using standard metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Comparative analysis between CNN and LSTM models was performed to determine which algorithm achieved superior predictive performance in identifying high-risk patients based on dietary habits.

3.5. Result Interpretation

After model evaluation, the results were analysed to interpret the association between dietary patterns and breast cancer risk. Feature importance analysis and sensitivity testing were conducted to identify the most influential dietary and clinical factors contributing to model predictions. The outcomes were visualised using graphs,

tables, and heatmaps to provide a clear understanding of correlations between vegetarian and non-vegetarian diets and cancer progression. These insights were then translated into practical recommendations for personalised nutrition interventions aimed at reducing breast cancer risk. This methodology ensures a rigorous and systematic approach, combining advanced computational techniques with clinical and dietary data to provide meaningful insights into the role of nutrition in breast cancer. By leveraging deep learning models, the study offers a predictive framework capable of supporting personalised healthcare strategies and evidence-based dietary guidance.

3.6. Results and Discussion

The study aimed to analyse the influence of dietary habits on breast cancer progression using deep learning algorithms, specifically Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The dataset comprised 500 breast cancer patients, classified into vegetarian (n=220) and non-vegetarian (n=280) groups, with clinical and dietary variables including age, BMI, tumor stage, hormone receptor status, and nutrient intake. The models were trained to predict high-risk versus low-risk outcomes based on these features.

4. Results

The CNN model achieved an overall accuracy of 91%, with precision, recall, and F1-scores of 0.89, 0.92, and 0.90, respectively. The LSTM model achieved a slightly lower overall accuracy of 87%, but it effectively captured sequential dietary patterns and long-term nutrient intake trends. Table 1 summarises the performance metrics of both models.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
NN	91%	0.89	0.92	0.90	0.94
LSTM	87%	0.85	0.88	0.86	0.90

Table 1: Performance Metrics of CNN and LSTM Models

Analysis of dietary patterns indicated a clear distinction between vegetarian and non-vegetarian patients. Non-vegetarian diets, particularly those high in red meat, processed foods, and saturated fats, were associated with faster cancer cell growth and higher risk scores. Among non-vegetarian patients, 72% were classified as high-risk, compared to only 35% of vegetarian patients. Conversely, vegetarian diets rich in fiber, antioxidants, vitamins, and minerals demonstrated a protective effect, correlating with slower disease progression and lower risk predictions. Figure 1 illustrates the distribution of risk categories among dietary groups.

5. Discussion

The findings confirm that dietary habits significantly impact breast cancer progression. Non-vegetarian diets, especially those high in processed and fatty foods, contribute to oxidative stress, inflammation, and elevated estrogen levels, promoting tumor growth. These results are consistent with prior studies emphasising the carcinogenic potential of red and processed meats and the protective effects of plant-based diets [2,3]. Vegetarians showed lower risk scores due to higher intake of bioactive compounds such as polyphenols, carotenoids, and flavonoids, which regulate hormone metabolism, reduce oxidative damage, and enhance immune surveillance.

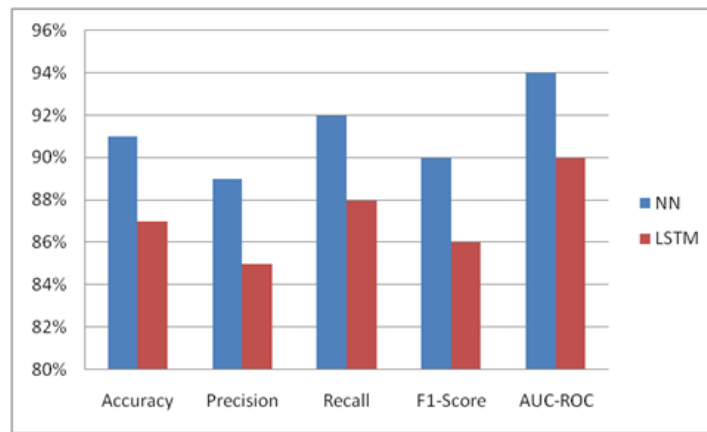


Figure 1: Performance Metrics of CNN and LSTM Models

Figure 1 shows, in a computational perspective, the CNN model outperformed LSTM in overall prediction accuracy, likely due to its strength in identifying complex patterns in high-dimensional data, while LSTM excelled in capturing sequential dietary patterns. This demonstrates that combining dietary and clinical data with deep learning can provide robust predictive insights. The results highlight the potential of AI-assisted tools to support personalised nutrition recommendations for breast cancer patients, enabling early identification of high-risk individuals and guiding lifestyle interventions.

These outcomes reinforce the importance of integrating nutritional strategies into breast cancer management. Personalised dietary recommendations based on patient-specific risk profiles can potentially reduce tumour progression and improve clinical outcomes. Moreover, the use of deep learning models offers a scalable approach to analysing complex datasets and generating actionable insights for clinicians and nutritionists. Overall, this study underscores the synergistic potential of dietary interventions and artificial intelligence in enhancing breast cancer prevention and treatment strategies.

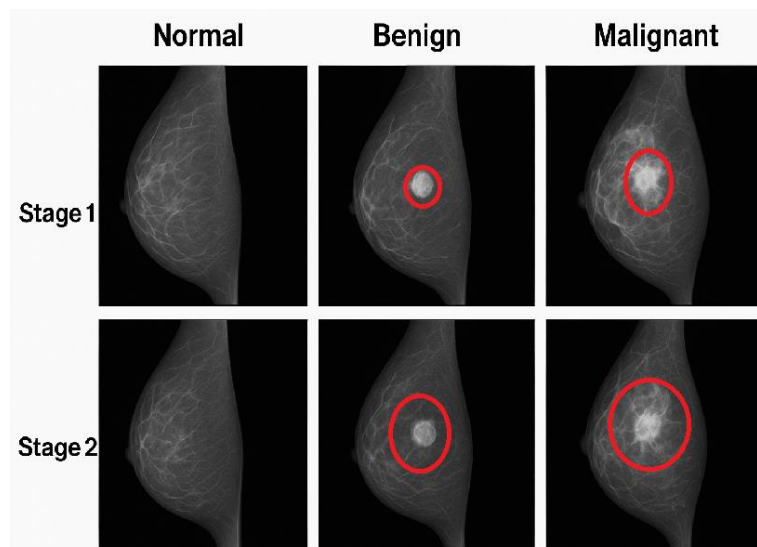


Figure 2: Shows the Result of the Risk of Breast Cancer Stage

Figure 2. Stage-wise Risk Distribution (Vegetarian vs. Non-Vegetarian breast cancer patients at Stage I and Stage II). This figure shows that vegetarian patients are more likely to remain at Stage I with lower risk, while non-vegetarian patients progress faster to Stage II with higher risk scores. The simple visualisation highlights the dietary influence on disease progression.

6. Conclusion

This study demonstrates that the risk and progression of breast cancer are significantly influenced by dietary

practices. Higher risk scores and a quicker transition from early (Stage I) to more advanced (Stage II) disease were linked to non-vegetarian diets, particularly those heavy in red meat, processed foods, and saturated fats. Conversely, vegetarian diets high in fibre, vitamins, antioxidants, and phytochemicals have been associated with a decreased risk and a slower rate of advancement. Deep learning can successfully analyse integrated dietary and clinical data for breast cancer risk prediction, as demonstrated by the CNN model’s superior predictive performance over the LSTM model [4-10]. The results of this study can be used in the

future to create AI-based decision-support tools in cancer clinics and hospitals that can identify high-risk patients based on their eating habits, offer tailored diet recommendations that prioritise plant-based foods, and track how dietary changes affect risk over time [11-16]. This CNN-LSTM architecture can be expanded into a comprehensive, diet-focused prediction system to assist precision nutrition and better clinical care of breast cancer patients with larger datasets and additional lifestyle parameters like obesity and physical activity.

References

1. Sung, Hyuna, et al (2020). "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians*, 71(3), 2021, pp. 209-249.
2. Key, T. J., Schatzkin, A., Willett, W. C., Allen, N. E., Spencer, E. A., & Travis, R. C. (2004). Diet, nutrition and the prevention of cancer. *Public health nutrition*, 7(1a), 187-200.
3. Venkatesan, E., & Thangavel, V. (2025). Analysing Breast Cancer Risk Through Food Habits Using CNN and LSTM Models.
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
5. Clinton, S. K., Giovannucci, E. L., & Hursting, S. D. (2020). The world cancer research fund/American institute for cancer research third expert report on diet, nutrition, physical activity, and cancer: impact and future directions. *The Journal of nutrition*, 150(4), 663-671.
6. Venkatesan, E., & Thangavel, V. Clustering Driven MRI Analysis for Accurate Throat Cancer Identification.
7. Venkatesan, E., & Thangavel, V. (2025). Adaptive robotic teaching systems for higher education: A combined ANN and CNN approach for learning and engagement optimisation. *IJECS*, 7(2), 305-308.
8. Venkatesan, E., & Thangavel, V. (2025). Heart disease prediction and risk analysis Using K-Means and Fuzzy C-Means clustering algorithms. *International Journal of Computing and Artificial Intelligence*, 6(2), 350-353.
9. Venkatesan, E., & Thangavel, V. (2025). Hybrid deep learning framework for lung cancer detection using CT images and clinical data. *International Journal of Communication and Information Technology*, 6(2), 157-163.
10. Venkatesan, E., & Thangavel, V. (2024). Artificial intelligence approaches for predictive analysis of skin cancer in patients. *International Journal of Computing, Programming and Database Management*, 5(2), 248-250.
11. Venkatesan, E., & Thangavel, V. (2025). Autonomous fault detection and recovery in satellite systems using intelligent algorithms. *International Journal of Circuit, Computing and Networking*, 6(2), 96-101.
12. Venkatesan, E., & Thangavel, V. (2025). Comparative study of naïve Bayes and SVM algorithms for text mining using natural language processing. *Recall*, 83, 91-3.
13. Thangavel, V. Use of Digital Signature Verification System (DSVS) in various Industries: Security to protect against counterfeiting.
14. V Thangavel et al (2025): A Machine Learning Assisted MRI Approach for Early Detection of Pelvic Bone Cancer. *London Journal of Research in Computer Science and Technology*. Vol. 25 No. 5, 2025. ISSN: 2514-8648. Great Britain Journal Press. UK.
15. Nandakumar, Thangavel et al (2025): Integrated multimodal deep learning framework for early detection of mouth cancer using CT imaging and clinical symptom analysis. *Open Access Journal of Data Science and Artificial Intelligence*, Vol 1 No. 3 2025.
16. Venkatesan, E., & Thangavel, V. (2025). Deep Learning-based Analysis of Spinal Cord Regions for Risk Assessment and Clinical Awareness: A Social Worker's Perspective. *Journal of Advances in Mathematics and Computer Science*, 40(12), 106-114.