Concept Paper

Leveraging ChatGPT for Enhancing Breast Cancer Detection Using Machine Learning

Shreyansh Singh *

Posted Date: 13 November 2024

doi: 10.20944/preprints202411.0953.v1

Keywords: Cancer

 doi:10.20944/preprints202411.0953.v1

Concept Paper

Leveraging ChatGPT for Enhancing Breast Cancer Detection Using Machine Learning

Shreyansh Kumar Singh [1,2]

[1] CSE Department, VIT University; sshreyansh822@gmail.com
[2] VIT Bhopal University, Bhopal, Madhya Pradesh, India

Abstract: Breast cancer is one of the most common and life-threatening diseases affecting women globally. Early and accurate detection is crucial for improving patient outcomes and reducing mortality rates. Machine learning (ML) models have demonstrated remarkable potential in assisting with breast cancer diagnosis through data-driven approaches that can identify patterns in complex datasets more effectively than traditional methods. In this context, ChatGPT, an advanced language model, offers valuable assistance by enhancing several aspects of the breast cancer detection process, such as model development, data preprocessing, and troubleshooting. It supports researchers by providing real-time code generation, model optimization techniques, and detailed explanations of ML concepts, thereby streamlining the process of building accurate and robust models. This paper discusses the ways ChatGPT aids in breast cancer detection using machine learning, including technical guidance, code generation, and knowledge enhancement, making AI-driven healthcare research more accessible.

Keywords: Cancer

I. Introduction

Breast cancer diagnosis typically involves the analysis of medical data, such as histopathological images or diagnostic datasets like the Breast Cancer Wisconsin (Diagnostic) Data Set. Machine learning techniques, including Support Vector Machines (SVM), Neural Networks, and Decision Trees, are often employed to identify patterns and make predictions based on this data. While domain expertise is essential for building such models, having access to AI-based assistance, like ChatGPT, can significantly expedite and improve the process.

II. Role of Machine Learning in Breast Cancer Detection

ML algorithms, such as SVM, Random Forest, and Convolutional Neural Networks (CNN), are commonly used in breast cancer detection. These models can identify malignancies with high accuracy after being trained on medical datasets. A crucial step in this process is preprocessing the data, which includes normalization, feature selection, and handling missing values. ML models also require fine-tuning through hyperparameter optimization to maximize their performance.

III. How ChatGPT assists in Breast Cancer Detection using M.L.

- **Model Development Guidance**: ChatGPT can provide suggestions and guidance on choosing appropriate ML models for breast cancer detection based on the dataset and the problem requirements. For example, it may recommend SVMs for structured data or CNNs for image-based classification tasks.
- **Code Generation and Troubleshooting**: ChatGPT can assist researchers and developers in writing efficient code for data preprocessing, model training, and evaluation. It can generate code snippets in Python using popular libraries such as scikit-learn, TensorFlow, and Keras, saving time and reducing the possibility of coding errors.

 doi:10.20944/preprints202411.0953.v1

- **Data Preprocessing Support**: Preprocessing medical datasets, especially in cases of missing values, imbalanced data, and scaling issues, is critical for building accurate models. ChatGPT can offer advice on the best preprocessing techniques, including normalization, scaling, and feature extraction, to ensure that the ML model receives clean, well-prepared data for training.
- **Hyperparameter Tuning**: Machine learning models require tuning of parameters such as the regularization factor in SVMs or the learning rate in neural networks. ChatGPT can provide step-by-step guidance on hyperparameter tuning, suggest best practices, and even offer code to implement grid search or random search methods.
- **Model Evaluation and Improvement**: ChatGPT can help in evaluating model performance by explaining various evaluation metrics like accuracy, precision, recall, and F1-score. It can also suggest strategies for improving model accuracy, such as cross-validation, feature engineering, and the use of ensemble methods.
- **Knowledge Enhancement**: ChatGPT can provide real-time explanations of complex ML concepts, offer insights into recent advancements in breast cancer detection techniques, and help in interpreting research papers. This makes it an invaluable tool for students, researchers, and professionals working on healthcare-related ML projects.

IV. Case Study: Breast Cancer Wisconsin (Diagnostic) Dataset

In a practical scenario, ChatGPT can assist in building an SVM model for breast cancer classification using the Breast Cancer Wisconsin (Diagnostic) dataset. After loading and normalizing the data, ChatGPT can guide users through the process of splitting the dataset, training the SVM model, tuning the hyperparameters, and evaluating the results, achieving high accuracy in predicting malignant and benign cases.

V. Conclusion

In a practical scenario, ChatGPT can assist in building an SVM model for breast cancer classification using the Breast Cancer Wisconsin (Diagnostic) dataset. After loading and normalizing the data, ChatGPT can guide users through the process of splitting the dataset, training the SVM model, tuning the hyperparameters, and evaluating the results, achieving high accuracy in predicting malignant and benign cases.

References

1. Wolberg, W. H., & Mangasarian, O. L. (1990). Mult surface method of pattern separation for medical diagnostics applied to breast fine-needle aspiration cytology. *Cancer Letters*, 77(2), 163-171. DOI: 10.1016/0304-3835(90)90155-4.
2. UCI Machine Learning Repository. (n.d.). Breast Cancer Wisconsin (Diagnostic
3. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. DOI: 10.1007/BF00994018.
4. Chawla, N. V., De Silva, V., & Matsumoto, K. (2005). An introduction to support vector machines. *Proceedings of the 2005 International Conference on Machine Learning and Applications*, 7, 25-29. DOI: 10.1109/ICMLA.2005.118.
5. Google Maps API. (n.d.). Google Maps Platform Documentation. Retrieved from https://developers.google.com/maps/documentation.
6. Brown, M. P., & Reinhold, W. C. (2002). A Systematic Review of the Diagnostic Accuracy of Clinical Breast Examination for Breast Cancer. *Journal of the National Cancer Institute*, 94(1), 1-5. DOI: 10.1093/jnci/94.1.1.
7. Sohail SS, Siddiqui J, Ali R. User feedback scoring and evaluation of a product recommendation system. In2014 seventh international conference on contemporary computing (ic3) 2014 Aug 7 (pp. 525-530). IEEE.
8. Farhat F, Sohail SS, Siddiqui F, Irshad RR, Madsen DØ. Curcumin in wound healing—a bibliometric analysis. Life. 2023 Jan 4;13(1):143.
9. Areeb QM, Nadeem M, Sohail SS, Imam R, Doctor F, Himeur Y, Hussain A, Amira A. Filter bubbles in recommender systems: Fact or fallacy—A systematic review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023 Nov;13(6):e1512.
10. Sohail SS, Siddiqui J, Ali R. Book recommender system using fuzzy linguistic quantifier and opinion mining. InIntelligent Systems Technologies and Applications 2016 2016 (pp. 573-583). Springer International Publishing.

 doi:10.20944/preprints202411.0953.v1

11. Sohail SS, Madsen DØ, Himeur Y, Ashraf M. Using ChatGPT to navigate ambivalent and contradictory research findings on artificial intelligence. Frontiers in Artificial Intelligence. 2023 Jul 27;6:1195797.
12. Alam MT, Ubaid S, Sohail SS, Nadeem M, Hussain S, Siddiqui J. Comparative analysis of machine learning based filtering techniques using MovieLens dataset. Procedia Computer Science. 2021 Jan 1;194:210-7.
13. Sohail SS, Siddiqui J, Ali R. A novel approach for book recommendation using fuzzy based aggregation. Indian Journal of Science and technology. 2017 May;8(1).
14. Muzaffar A, Nafis MT, Sohail SS. Neutrosophy logic and its classification: an overview. Neutrosophic Sets and Systems. 2020 Sep 4;35:239-51.
15. Irshad RR, Hussain S, Sohail SS, Zamani AS, Madsen DØ, Alattab AA, Ahmed AA, Norain KA, Alsaiari OA. A novel IoT-enabled healthcare monitoring framework and improved grey wolf optimization algorithm-based deep convolution neural network model for early diagnosis of lung cancer. Sensors. 2023 Mar 8;23(6):2932.
16. Sohail SS, Khan MM, Arsalan M, Khan A, Siddiqui J, Hasan SH, Alam MA. Crawling Twitter data through API: A technical/legal perspective. arXiv preprint arXiv:2105.10724. 2021 May 22.
17. Farhat F, Silva ES, Hassani H, Madsen DØ, Sohail SS, Himeur Y, Alam MA, Zafar A. The scholarly footprint of ChatGPT: a bibliometric analysis of the early outbreak phase. Frontiers in Artificial Intelligence. 2024 Jan 5;6:1270749.
18. Sohail SS, Siddiqui J, Ali R. Classifications of recommender systems: A review. Journal of Engineering Science & Technology Review. 2017 Jul 1;10(4).
19. Alsagri HS, Sohail SS. Fractal-Inspired Sentiment Analysis: Evaluation of Large Language Models and Deep Learning Methods. Fractals. 2024 Aug 30.