INNOVATIVE PUBLICATION
Journal of Effective Learning and Sustainable Innovation
Vol.3 №10 (2025). October
innovativepublication.uz

Development of an Algorithm and Software Tool for Early Diagnosis of Breast Cancer Based on Intelligent Analysis

Mukhamediyeva Dilnoza,
Professor of "TIIAME" NRU,
d.muhamediyeva@tiiame.uz
Khamraev Mansur,
Student of TUIT FSE, *mxamrayev888@gmail.com*

Annotation

Breast cancer is one of the most common and life-threatening diseases among women worldwide. Early diagnosis plays a crucial role in increasing survival rates and improving treatment effectiveness. This article presents an approach that leverages artificial intelligence (AI) and machine learning (ML) techniques to develop an algorithm and software tool for the early detection of breast cancer. The proposed system integrates medical imaging, deep learning models, and data-driven analytics to enhance diagnostic accuracy.

Keywords

Breast cancer detection, Artificial Intelligence, Machine Learning, Deep Learning, Medical Imaging, Convolutional Neural Networks, Support Vector Machines, Random Forest, K-Nearest Neighbors, Artificial Neural Networks, Recurrent Neural Networks, Feature Extraction, Early Diagnosis.

Introduction

Breast cancer detection at an early stage significantly increases the chances of successful treatment. Traditional diagnostic methods, such as mammography and biopsy, have limitations in terms of accuracy and accessibility. With the rapid advancements in AI, computational techniques can assist healthcare professionals by providing reliable and automated diagnostic support.

Methodology

The proposed system is designed to process and analyze medical imaging data using deep learning algorithms. The key components of the methodology include:

1. **Data Collection**: The dataset consists of mammographic images, patient history, and biopsy results obtained from medical institutions.
2. **Preprocessing**: Image enhancement techniques, such as noise reduction and contrast adjustment, are applied to improve image quality.

3. **Feature Extraction**: Convolutional Neural Networks (CNNs) are used to extract meaningful features from images.
4. **Classification Models**: Several machine learning and deep learning models are utilized to improve accuracy, including:
- **Convolutional Neural Networks (CNNs)**: Used for image analysis and feature extraction.
- **Support Vector Machines (SVMs)**: Applied for classifying benign and malignant tumors based on extracted features.
- **Random Forest (RF)**: A decision-tree-based ensemble learning technique for pattern recognition.
- **K-Nearest Neighbors (KNN)**: Used for comparative classification based on similarity metrics.
- **Artificial Neural Networks (ANNs)**: Applied for learning complex patterns in patient data and imaging.
- **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)**: Used for analyzing sequential patient history data.

5. **Software Implementation**: A user-friendly software tool is developed to integrate the trained models, allowing medical professionals to upload images and receive diagnostic results with confidence scores.
6. **Patient Features Used**: The models consider various patient-related features, including:
- **Tumor Size**: The physical dimensions of the detected tumor.
- **Cell Shape and Texture**: Characteristics of tumor cells, such as smoothness, concavity, and uniformity.
- **Clump Thickness**: A measure of the tumor cell density.
- **Mitotic Count**: The rate of cell division, which is an indicator of malignancy.
- **Hormone Receptor Status**: Presence of estrogen and progesterone receptors.
- **Age and Medical History**: Patient age and history of breast cancer or related conditions.
- **Genetic Mutations**: BRCA1 and BRCA2 gene mutations if available.
- **Lymph Node Involvement**: Presence of cancerous cells in lymph nodes.

7. **KNN Algorithm Calculation Process**: The K-Nearest Neighbors (KNN) algorithm follows these steps:
- **Step 1: Data Preparation**: The dataset is split into training and testing sets, where each sample is represented as a feature vector.

- **Step 2: Distance Calculation**: The Euclidean distance between the test point and all training samples is computed using the formula: where and are feature vectors, and is the number of features.
- **Step 3: Selecting Neighbors**: The k-nearest data points (typically an odd number) are identified based on the shortest computed distances.
- **Step 4: Majority Voting**: The class labels of the selected neighbors are examined, and the most frequent label is assigned to the test sample.
- **Step 5: Classification Output**: The predicted class (benign or malignant) is provided based on the majority vote.

Example: Suppose we have a new patient with the following features: Tumor size: 1.2 cm, Clump thickness: 4, Cell texture: 7, mitotic count: 2, Age: 45.

The KNN algorithm calculates the distance of this sample from previously labeled cases in the dataset. If k = 3, and the three nearest neighbors belong to the following classes: Benign (Distance: 0.8), Malignant (Distance: 0.9), Malignant (Distance: 1.0)

Since the majority of the three neighbors are malignant, the new sample is classified as **malignant**.

Results and Discussion

The algorithm is evaluated using performance metrics such as accuracy, sensitivity, specificity, and F1-score. The performance results of different models are as follows:

- **CNN**: Achieved an accuracy of 95%, with a high sensitivity of 96% in detecting malignant cases.
- **SVM**: Provided an accuracy of 91% but had a lower sensitivity of 89% compared to deep learning approaches.
- **Random Forest**: Showed an accuracy of 90%, performing well in handling noisy datasets but slightly lower in sensitivity at 87%.
- **KNN**: Performed with an accuracy of 88%, but was slower in classification due to high computational requirements.
- **ANN**: Achieved an accuracy of 93%, effectively learning complex relationships in medical data.
- **RNN/LSTM**: Best suited for analyzing patient history, achieving an accuracy of 92% when integrating sequential data with imaging results.

These results were obtained from publicly available datasets such as the Wisconsin Breast Cancer Dataset (WBCD) and the Digital Database for Screening Mammography (DDSM). Additionally, experimental evaluations were conducted using Python-based AI frameworks like TensorFlow and PyTorch, leveraging real-world clinical datasets from medical institutions.

Preliminary tests indicate that the proposed model outperforms traditional diagnostic approaches, reducing false positives and false negatives. The software tool provides real-time analysis and can be integrated with existing hospital management systems.

Conclusion

Intelligent analysis and AI-based diagnostic tools have the potential to revolutionize breast cancer detection. The developed algorithm and software tool contribute to early diagnosis, improving patient outcomes and reducing diagnostic errors. Future enhancements include integrating multimodal data sources, improving model robustness, and expanding the dataset for better generalization.

References

1. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. CA Cancer J Clin. 2022.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021.
3. Taylor C, McGale P, Probert J, Broggio J, Charman J, Darby SC, et al. Breast cancer mortality in 500 000 women with early invasive breast cancer diagnosed in England, 1993–2015: population based observational cohort study. BMJ. 2023.
4. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013.
5. The Royal College of Radiologists. RCR Clinical Radiology Workforce Census 2022. London: The Royal College of Radiologists; 2022.