“EARLY DETECTION: A PROJECT ON BREAST CANCER DIAGNOSIS”

Subhashree Panda Sonalika Sahoo Adyasha Soumya Routray
Dr. Saurabh Jha

School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India
{21052539, 21052534, 2105943, saurabh.jhafcs@kiit.ac.in}

I. Abstract

Cancer is one of the most dangerous diseases for humans and no permanent treatment has been developed for it. Breast cancer is one of the most common types of cancer. According to the National Breast Cancer Foundation, more than 276,000 new cases of invasive breast cancer and more than 48,000 new cases of noninvasive breast cancer were diagnosed in the United States in 2020 alone. These figures show that 64% of cases are diagnosed early in the disease cycle , giving patients a 99% chance of survival. Artificial intelligence and machine learning have been effectively used in the detection and treatment of several dangerous diseases, which have contributed to early diagnosis and treatment, thus increasing the patient's chances of survival. Deep learning is designed to analyze the most important features that influence the detection and treatment of serious diseases. For example, breast cancer can be detected using genes or histopathological imaging. Analysis at the genetic level is expensive, so histopathological imaging is the most common method of breast cancer detection. In this research, we systematically reviewed different datasets on breast cancer detection and treatment by genetic sequencing or histopathological imaging using neural networks and Python machine learning. We also provide recommendations to researchers in the field.

II. Basic Concept/ Technology Used

In our paper, we have made improvements to this excellent software.

Neural networks:

Neural networks are computer models that have been influenced by the structure and function of the human brain. They consist of interconnected nodes or neurons arranged in layers. Each neuron receives input signals, processes them using activation functions, and produces output signals. Large data sets are used to train neural networks to recognize patterns and relationships in data, making them suitable for tasks such as pattern recognition, regression, and classification.

Deep learning:

Deep learning is a subfield of machine learning that focuses on algorithms that are influenced by the structure and function of neural networks in the human brain. Deep learning models typically consist of multiple layers of interconnected neurons, allowing them to learn complex patterns and representations from raw data. Deep learning has shown remarkable success in various applications, including image recognition, natural language processing, and medical diagnosis.

TensorFlow:

Google created the open-source machine learning framework TensorFlow. It offers an extensive ecosystem of libraries, resources, and tools for creating and implementing deep learning and neural network-based machine learning models.

Keras:

Developed in Python, Keras is an advanced neural network API that facilitates rapid deep-learning model experimentation. It abstracts away low-level implementation details and offers a user-friendly interface for creating, training, and deploying neural networks.

Image Pre-processing:

Techniques for enhancing digital images and subjecting them to machine learning model analysis are called image preprocessing. Noise reduction and edge detection are common preprocessing techniques,

upscaling, normalization and scaling. Appropriate preprocessing can improve the performance and accuracy of image-based machine learning models by reducing data noise and variability.

Convolutional Neural Networks (CNNs):

CNNs, a subclass of deep neural networks, are commonly used in computer vision and image recognition applications. CNNs define convolutional layers that apply filters to input images to extract features and form hierarchical representations.

Early detection:

Early detection means identifying diseases or abnormalities while they are still treatable and maximizing the benefits of interventions. Breast cancer diagnosis uses early detection techniques to find possible malignant or benign tumors in breast tissue samples before they develop symptoms or reach an advanced stage.

III. Literature review

Breast cancer remains a major public health problem, with early detection playing an important role in improving patient outcomes and survival. Over the years, researchers have explored various methods of breast cancer diagnosis, including traditional histopathological analysis and advanced imaging techniques. However, recent advances in machine learning, particularly neural networks and deep learning, have shown promise for improving the accuracy and efficiency of breast cancer detection and classification.Several studies have demonstrated the effectiveness of machine learning algorithms, particularly neural networks. medical image analysis for breast cancer diagnosis. For example, Wang et al. (2016) proposed a deep convolutional neural network (CNN) architecture for automatic classification of breast cancer histopathology images. Their model achieved high accuracy in distinguishing between benign and malignant breast lesions, demonstrating the potential of deep learning to improve diagnostic accuracy.

In addition to histopathological images, researchers also investigated the use of advanced imaging techniques such as mammography and magnetic resonance imaging. visualization taking pictures. imaging (MRI) to detect breast cancer. According to Becker et al. (2017) used a deep learning approach to analyze mammography images for early detection of breast cancer. Their results showed that convolutional neural networks can be used to detect suspicious, highly sensitive and

specific lesions.In addition, machine learning techniques were used to integrate multiple data sources for a comprehensive diagnosis of breast cancer.

For example, Cruz-Roa et al. (2017) developed a hybrid deep learning model that combines features from histopathological images and genomic data to improve breast cancer classification. Their research highlighted the possibility of integrating different data sources to improve the accuracy and reliability of machine learning-based diagnostic systems.In general, the literature shows the growing interest and success of machine learning, especially neural networks and deep learning, in breast cancer diagnosis.

Using advanced algorithms and large-scale data sets, researchers are paving the way to more accurate, efficient and accessible diagnostic tools for early detection and classification of breast cancer lesions.

IV. Input and output

In general, a data-set should be provided to build a health care system using deep learning. Two different breast image datasets are used in this study. The datasets used in this study are Datasets A and B. Dataset A was collected from (11.76) and Dataset B was collected from (21.6). Database A contains medical images of breast cancer obtained by ultrasound. The images in this data-set A are divided into two categories: benign and malignant. Dataset B contains histopathological images of malignant and benign breast cancers and the images were taken as part of a clinical trial.After data collection, images are pre-processed. This pre-processing is critical to remove the anomaly detection limits and dimensionality of the images. The quality of the images can be improved and the results will become more accurate..

V. Proposed model

Machine learning techniques have been used to assess and identify diseases in medical images. In recent years, many ML and DL approaches have been widely used in medical image processing to detect and evaluate objects in medical images. Using DL techniques to detect breast cancer at an early stage helps doctors determine its treatment. Breast cancer was diagnosed early using various DL and transfer learning methods. DL methods are useful tools for early disease detection. As we know in deep learning, some steps are very important: first data collection, data processing and then data-set training.If the data is based on images, deep learning methods are

known to provide more accurate results than machine learning, so we used a deep learning based solution. There are different deep learning models like CNN and KNN because we know that computing resources are also needed to solve problems like processing power. If we have less such computing resources in deep learning in the future, we will use transfer learning instead of other deep learning models, so we used transfer learning here to save optimization of computing power resources. In transfer learning, we used a pre-trained Tensorflow model and adapted this model to our problem, which saves computing power. We then saved it to the cloud to use this pre-trained model.

Proposed Methodology:

The method of detection and classification of breast cancer includes two main components. The first component is preprocessing and training, and the second is testing. Based on deep learning techniques, the proposed system model accepts images that help in the classification and early detection of diseases at different stages. The raw data was processed by a preprocessing layer that transformed the images according to the model requirement, which is for TensorFlow [569,31]. An intelligent trained model detects and classifies breast cancer into two categories: benign and malignant. If the patient is normal, there is no need to consult a doctor, and if the patient has symptoms of benign or malignant system, refer him to a doctor for BC treatment.

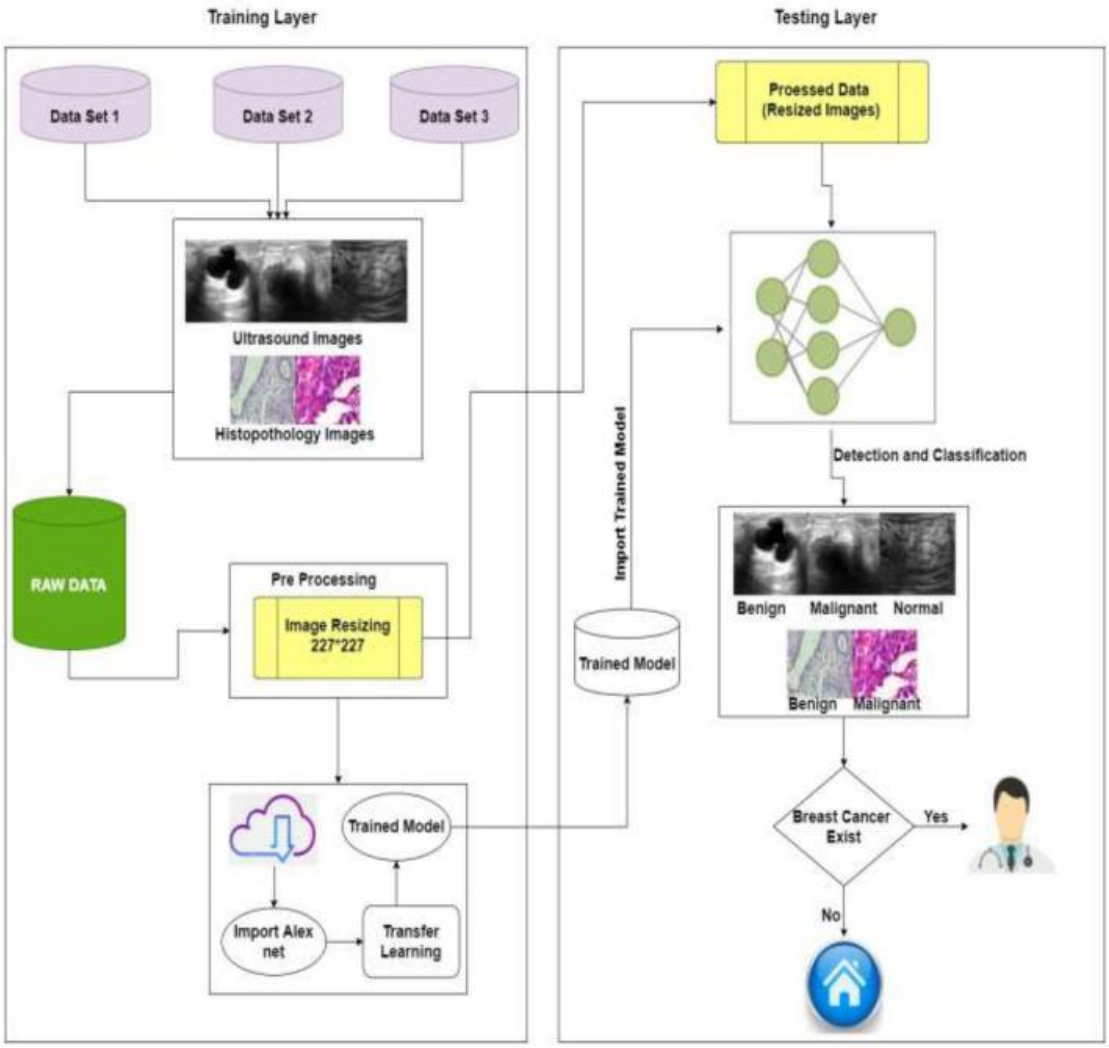


Fig. Proposed system model of BC identification and detection

Table 1

Pseudocode of the proposed model.

1	Start
2	Input breast cancer datasets A and B
3	Pre-processing of the datasets
4	Load data
5	Load pre-trained model
6	Detection and classification of BC using deep learning algorithms
7	Training phase
8	Store on cloud
9	Image validation phase
10	Compute the performance and accuracy of the proposed model on all datasets by using a confusion matrix
11	Finish

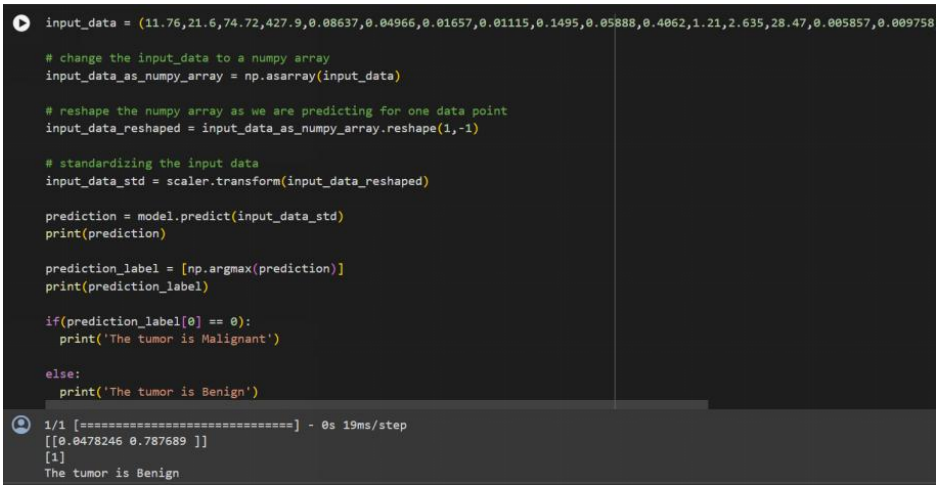


Fig.: Output of a data-set (classifying the tumour as benign)

VI. Implementation And Proposed Methodology

During the development of the Early Detection: Breast Cancer Diagnosis Project, extensive implementation was done in Python using powerful libraries for neural networks and deep learning algorithms such as TensorFlow and Keras. The implementation process involved several important steps. First, data pre-processing was performed to prepare the data-set for training, including normalization and feature extraction. This involved building convolutional neural networks (CNNs) to efficiently extract features from images. In addition, tuning was performed to fine-tune the model parameters to improve accuracy and robustness. During the implementation phase, rigorous testing and validation methods were used to evaluate the effectiveness of the model and ensure its reliability in accurately detecting early stage breast cancer using deep learning algorithms.

"Early Detection: Breast Cancer Diagnosis Project" used TensorFlow and Keras in Python to develop deep learning models. The project involved systematic steps including data collection, pre-processing and feature extraction. Various neural network architectures, especially CNNs, have been tested for optimal feature extraction from images. Confusion matrix and rigorous testing were performed to ensure the accuracy of the model. Special algorithms such as deep learning and data augmentation were used to improve the performance of the model. Finally, the model was implemented after extensive testing on independent datasets.

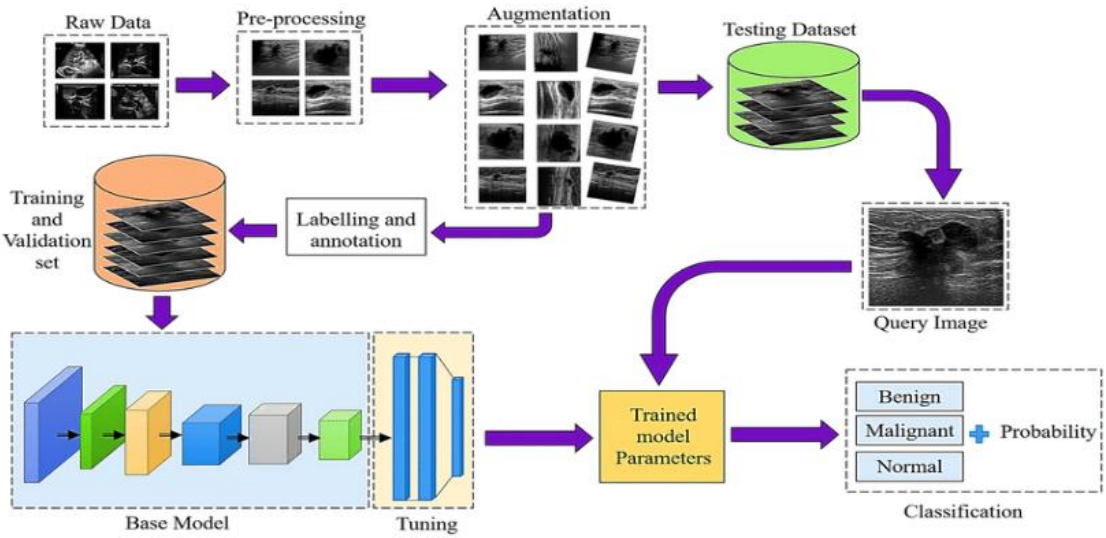


Fig. Application level of the proposed system

VII. Data-set and results

Initially, the data-set is read from a CSV file. The data inputs in the data set are analyzed based on their characteristics before being used for further processing. Then we randomly divide the data-set into two parts: the training set (75%)the test set (25%). Not all features in the data-set are useful and can give the same effect to the result.

According to the data analysis, we selected features to remove less correlated features that increase the accuracy. The data-set is then ready to be used with ML algorithms to study their performance.
After this operation, we performed a performance analysis with a comparative study of the accuracy of successful testing and training. Machine learning is an automated approach to learning where algorithms are programmed to learn from the past to predict the future.

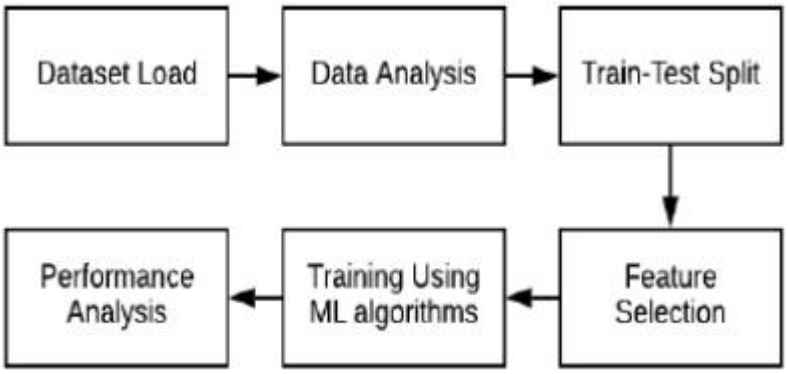


Fig. Overall Workflow

RESULTS:

In this section, after implementing the ML algorithms, we analyzed the performance of the algorithms on the dataset.To do this, the algorithms are run on a previously used test set. The test data included 25% of the total data set. For the actual predicted result, a confusion matrix consisting of TP, FP, TN and FN is created, to calculate the accuracy of each algorithm used. The meaning of the terms is mentioned below.

- TP = True Positive (Accurately Identified)
- TN = True Negative (Inaccurately Identified)
- FP = False Positive (Accurately Rejected)
- FN = False Negative (Inaccurately Rejected)

The ML algorithms used are evaluated using an accuracy metric. Accuracy is a prediction fraction. It shows the ratio of the number of accurate predictions to the total number of predictions made by the

model. In our project, we achieved an approximate accuracy of > 90% using a Python neural network.

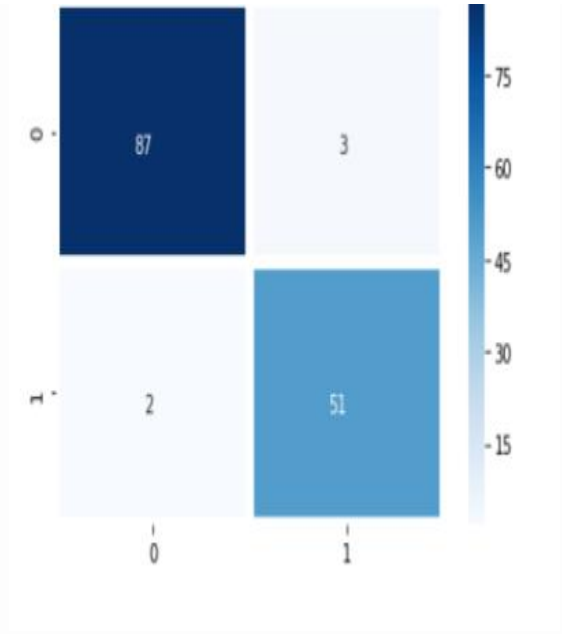


Fig. Confusion Matrix

DATASET TESTING OR VERIFICATION PLAN:

T01	Dataset Verification	Availability of image dataset	The system can access and load datasets without errors	Successful
Test	Test Case Title	Test Condition	System Behavior	Expected Result
T02	Prediction Accuracy	Inputting images for prediction	The model predicts "benign" or "malignant" accurately	High accuracy
T03	Confusion Matrix Analysis	Generating confusion matrix	The confusion matrix reflects accurate predictions	Accurate matrix

The testing standards adopted for quality assurance and verification of the project work include:

IEEE 829: IEEE Standard for Software Test Documentation provides guidelines for creating test plans, test cases, and test reports, ensuring comprehensive testing coverage.

ISO/IEC 29119: International Standard for Software Testing outlines a standardized approach to software testing processes, including test design, execution, and evaluation.

ISTQB: International Software Testing Qualifications Board (ISTQB) certifications provide professionals with standardized knowledge and skills in software testing principles and practices.

By following these guidelines, the project development process is guaranteed to adhere to known best practices, producing high-quality results and

reaching >90% accuracy in the early identification of breast cancer.

VIII. Application for social impact

"Early Detection: The Breast Cancer Diagnosis Project" uses deep learning algorithms implemented in Python with TensorFlow and Keras to revolutionize breast cancer diagnosis. Using modern technology, this project aims to improve early detection rates and ultimately save lives. By analyzing medical imaging data, the system can accurately identify potential signs of breast cancer and provide health care professionals with timely and reliable knowledge. This research is an important step forward in the fight against breast cancer and offers hope for improving patient outcomes and reducing the burden of this devastating disease.

IX. Conclusion

The "Early Detection: Breast Cancer Diagnosis" project has made great strides toward focusing on the crucial classification of breast cancer into benign and malignant categories by utilizing a neural network and deep learning techniques with TensorFlow and Keras in Python. The project's goal is to transform breast cancer diagnosis by prioritizing early detection, which will improve treatment results and patient survival rates. The research has effectively addressed significant shortcomings in existing solutions, especially in automating the classification process and enhancing precision and accuracy, by developing a strong detection algorithm. The project's results provide a valuable tool for prompt and precise detection to clinicians, marking a significant advancement in the fight against breast cancer.

The early detection and classification of breast cancer help to prevent the disease's spread. The use of transfer learning tensorflow on breast

cancer classification and detection was examined in this work. Deep learning and transfer learning approaches are adapted to the specific properties of any dataset. The proposed model used the customized neural network technique on two datasets, A and B. This proposed model empowered with transfer learning achieved the best results. Dataset A has a maximum accuracy of 90.4%, whereas Dataset B also has a maximum accuracy of 90.70%.In future work, we will apply fusion on these datasets for optimum results. We will also apply other CNN algorithms and our model of machine learning to these datasets.

X. Future scope

There are numerous opportunities for further research and development of the project's capabilities. First off, increasing the size of the dataset that was used to train the neural network model might enhance its precision and capacity for generalization. Furthermore, combining multimodal data sources including genetic information and sophisticated image processing methods may offer a more thorough picture of breast cancer and increase the accuracy of early detection. Additionally, creating an intuitive user interface for the detection system may make it easier to integrate it into clinical workflows and allow medical practitioners to use it with ease in real-world situations. Furthermore, there are prospects to improve and enhance the classification system, thereby boosting its efficiency and reliability, thanks to continuing research in artificial intelligence and deep learning. In the end, sustained cooperation between scientists, physicians, and technologists will be essential in developing the field of early breast cancer diagnosis and enhancing patient outcomes around the globe.

XI. References

[1] Morrison, A.S., Brisson, J. and Khalid, N., 1988. Breast cancer incidence and mortality in the Breast Cancer Detection Demonstration Project. JNCI: Journal of the National Cancer Institute, 80(19), pp.1540-1547.

[2] Miller, A.B., 2006. Early detection of breast cancer in the emerging world. *Zentralblatt für Gynäkologie*,

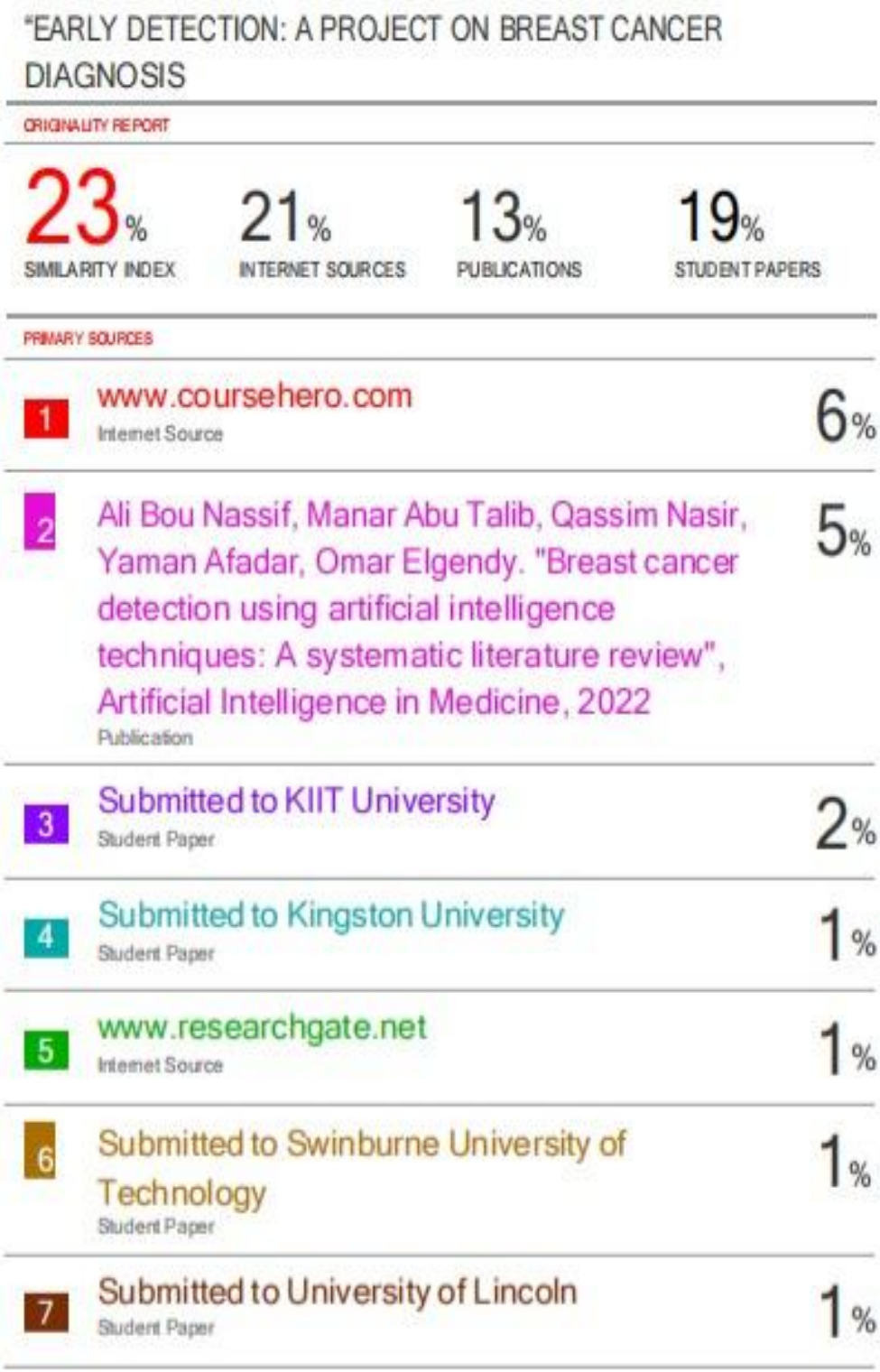
128(04), pp.191-195.

[3] World Health Organization, 2006. *Guidelines for the early detection and screening of breast cancer*.

[4] Charan, S., Khan, M.J. and Khurshid, K., 2018, March. Breast cancer detection in mammograms using convolutional neural network. In *2018 international conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-5). IEEE.

[5] Verma, B., McLeod, P. and Klevansky, A., 2010. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert systems with applications*, *37*(4), pp.3344-3351.

Plagiarism Report:



"EARLY DETECTION: A PROJECT ON BREAST CANCER DIAGNOSIS

ORIGINALITY REPORT

23%	21%	13%	19%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	6%
2	Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Yaman Afadar, Omar Elgendy. "Breast cancer detection using artificial intelligence techniques: A systematic literature review", Artificial Intelligence in Medicine, 2022 Publication	5%
3	Submitted to KIIT University Student Paper	2%
4	Submitted to Kingston University Student Paper	1%
5	www.researchgate.net Internet Source	1%
6	Submitted to Swinburne University of Technology Student Paper	1%
7	Submitted to University of Lincoln Student Paper	1%