iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71033

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Breast Cancer Detection Using CNN

Dr. Yeresime Suresh[1], Lakshmi Priya Naidu T[2], Lakshmi Srujana Sunkara[3], Mangala Gowri[4], Meeraj Mathin[5]
Dept. of CSE - Artificial Intelligence, Ballari Institute of Technology and Management, Ballari, India

Abstract: The article proposes an approach for detecting breast cancer with the aid of machine learning. The main motive of this article is to detect cancer at earlier stage using deep learning (DL) where the ultrasound images of breast are classified as cancerous or non-cancerous. Using CNN model the system is trained to classify based on tumour size. These ultrasound images are pre-processed using normalization and are resized beforehand so that the learned model can make predictions with ease. The system even detects the stage of cancer if the tumour is predicted to be cancerous using prediction probability. Overall this helps the medical professionals to detect breast cancer easily and provide proper medication to the affected ones and prevent human error.

Index Terms: Breast Cancer Detection, Machine Learning, Deep Learning, CNN, Tumor Classification, Early Detection, Ultrasound Images.

I. INTRODUCTION

Cancer is one of the major cause of death in the world today. World Health Organization(WHO) declared that the breast cancer accounts to nearly 25% of the diseases detected overall in the world today, where it goes on increasing every year. Its existence in humans explains how important is taking precautions and detecting the cancer early. Some methods to detect the cancer include mammography, biopsy, ultrasound imaging, and clinical examination. These methods cured the illness in quite several cases with an good number. However, mammography challenges still exist in few females who are having high tissue density in them because these cases do not notice the cancer in the human beings. The results of mammography are examined by the experienced radiologists, which leads to misdiagnosis or missed cases and, more unnecessary treatments or delayed diagnosis may happen. Detecting the breast cancer as early as possible could play a vital role as it may directly affect the treatment itself and it could also affect the future of the patient. goes straight into affecting the treatment itself and the future of the patient. If cancer is identified in the early stages, it helps in precisely handling the case & the probability of patient survive-ability increases. At this juncture, medical image processing is the distinct procedure enabling accurate detection of breast cancer, further enhancing improved diagnostic outcomes. Present scenarios are very effective but are time consuming, involve higher cost along with associated patient discomfort, with an additional burden of labor intensive process. Also human lapses are observed because of radiologists fatigue or biased judgment. Deep Learning (DL) methods provide powerful automation and efficient techniques for analysing complex medical images. The CNNs are better at detecting minute textures and patterns in ultrasound images which aid us in making more improved diagnostic decisions. Significant reduction can be observed w.r.t human intervention using machine learningbased automated analysis of images, thus mitigating misdiagnosis.

The main advantage of ultrasound detection of breast cancer is its use as a painless method, relatively inexpensive, and relatively accessible imaging tool. It is a method mainly useful for differentiating solid tumours from cysts. It has an excellent possibility in diagnosing breast cancers from patients having dense breast tissues such that the use of mammography would be invalid. However, as efficient, ultrasonography maintains skilled radiologists to make proper decisions and so may widen diagnostic differences. An automated system has been developed in this project which can successfully classify between malignant(cancerous) and benign(non-cancerous) tumour with the help of machine learning(ML) algorithms. An extensive labelled database has been used in order to train the model which help ML to recognise different shapes and sizes of tumour cells and understand the pattern accordingly. This way the system serves an great tool for the clinicians in order to diagnose breast cancer in the early stages itself.

This system automates the process of classifying the ultrasound images on the basis shape and size of breast tumour cells which makes this overall process of detection much faster, precise and are less likely to make wrong predictions followed by unnecessary treatment. This also makes it more convenient for radiologists and the rest of health practitioners to pay attention to other matters relative to the treatment of the patient. Generally for complicated cases there is an huge requirement of highly advanced equipment and skilled radiologists. But with the use of this automated system diagnose can be made easily with help of less advanced machines and it also reduces the requirement of highly skilled medical professionals.

The general motive behind developing this model is to make use of deep learning(DL) and machine learning(ML) methodologies and classify the tumorous cells much faster and accurately. The system will also meet the clinical demands of real-world applications by processing high volumes of ultrasound images, since time, accuracy, and efficiency are very essential in real- world clinical applications.

Apart from this, machine learning based detection approaches help in eliminating human errors and frequent mistakes caused by fatigue. Also this implies more consistency and reliability. Thus ensuring relatively early detection and proper planning can be carried out for treatment procedure. This approach will only come out with best survival rates by minimizing pressure on medical system, subsequently reducing treatment procedure also.

Finally, CNNs provide a new breakthrough towards the detection mechanism of breast cancer, as the system is able to make classification of any type of breast cancer. This augments faster and accurate detection for medical professionals. Also, contributing significantly in minimizing patients troubles with speedy recovery. The system undergoes significant complex testing continuously to improve feasibility, economic possibility and to be more reliable. This provides an great opportunity to increase the survival rate of a person suffering from breast cancer or at least that person can be treated with proper medication. The arrangement of rest of the paper is as follows: Section II examines the investigations that have been performed in connection with the intended task. Section III offers a concise an overview of the planned approach work. Section IV presents the attained results and discussions, and Section V serves as the conclusion for the study.

II. LITERATURE SURVEY

There has been an remarkable progress in the last twenty to thirty years in identifying cancer with the help of normalization of images and many more techniques that come under ML and DL. Emergence of these technologies have brought an significant change in the way the professionals diagnose and classify breast tumours. With increased accuracy and ability to diagnose quickly has decreased the dependence on hands-on interpretation of images.

The most ancient technique used in order to diagnose and cure breast cancer is mammography. In the past, many researchers have found and have focused their research to enhance the correctness of mammogram image analysis by incorporating models of machine learning.

Although mammography remains the gold standard in the detecting the cancer, ultrasound imaging has attracted widespread attention over time because it is non- invasive and cost-effective. Many studies demonstrate the potential of ML algorithms in ultrasound image analysis. Mazurowski et al. (2011) developed a method for classifying breast tumors as benign or malignant based on ultrasound images using texture features that were taken from the images. They used SVM and obtained promising classification accuracy. However, their approach was high in extracting feature intensive, which is time- consuming and computationally expensive.

Other such seminal contribution of Cireşan et al. during 2013 related to the deep learning idea used with mammography images toward designing algorithms in classifying a tumor. Deep CNN architecture for designing their model led to far greater accuracy levels compared to machine learning's conventional algorithm. However, these methods depend on large data for training purposes and, needless to say, are drawn processes.To keep up with these expectations, certain issues like limitations in the mammography-technology (breast tissues with density) are also stated along with it.

DL techniques, specially CNNs, have also been applied broadly for ultrasound image analysis. Zhang et al. (2018) proposed a DL-based model for the classification of breast tumors in ultrasound images. The model used a multi-layer CNN that automatically learned features from ultrasound images to eliminate the process of feature extraction manually. In their approach, they were very successful and established the fact that deep learning works very well by improving the breast cancer diagnosis with ultrasound images. This study also took the size and quality of dataset in training deep learning models and their potential applications in real-world clinical settings. Combining MachineLearning(ML) and Deep-Learning(DL) techniques with ultrasound imaging presents a promising opportunity for breast cancer identification, specially in patients whose breast tissue is dense and in whom mammography may not be effective.

Dey et al. 2019 and Morkel et al. (2018), have applied hybrid models that comprise several ML and DL algorithms in building more robust frameworks for the identification of breast cancer. Hybrid models are built by taking a composite of the classical ML algorithms like SVM, k-NN, etc. and the deep learning model in the form of CNNs to obtain more accuracy but with decreased false positives and negatives.

Rajpoot et al. in 2019 developed a deep learning-based system that classifies mammography images as either benign or malignant by applying Convolutional Neural Networks.

Their approach does achieve high sensitivity and specificity toward benign vs. malignant differentiation, but the model in practice hinges on quality images and requires preprocessing, including enhancement and noise removal.

In addition to this, many research works have focused more on the feature selection and optimization techniques to improve the working of the models. Tan et al. (2020) developed a method that improved the feature selection process by integrating both domain knowledge and automated feature selection algorithms into a hybrid approach. The hybrid approach thus identified the most important features for exact tumor classification and reduced the computational complexity of the model.

Much has been achieved in ML and DL in detecting of breast cancer. However, such large, well-labeled datasets remain a challenge to train models upon; the limitation of model interpretability and real-time applicability into the clinical scenario still applies. Further still, these systems need to be integrated into clinical workflows by collaborating computer scientists and medical professionals in making sure that the systems will be user-friendly and easily adoptable to healthcare providers.

In general, the related work in this area of breast cancer detection proves that ML and DL methods improve the correctness of diagnosis and reduce the burden on healthcare professionals. Further research is needed for the improving the systems that have these problems and flaws, and for making models more scalable and real-time applicable in the clinical environment.

This approach is designed to add such initiatives on the growth of automated ultrasonic imaging- and machinelearning-based breast cancer detection. It comes with much-needed accuracy in a fast setting suitable for clinical applications.

III. PROPOSED SYSTEM

This section highlights on the techniques of ML and DL used towards automatic detection in breast cancer w.r.t ultrasound images. Further accurate classification of tumor types- benign or malignant, which would ultimately be helpful to health care physicians to decide a line of action. Thus, the approach developed is an attempt to prepare this user- friendly tool, efficient enough to carry out the operation of handling lots of ultrasound images taken at clinical situations.

A. System Overview

1) Data Input:

The ultrasound images of the breast will be the primary data input to the system. Data can be captured by the special equipment or given to the system in digital form.

2) Pre-processing:

Changing the size of ultrasound images, normalization of ultrasound image and reducing the noise were all done to get ready for further analysis.

3) Classification:

It will classify the tumor to be benign or malignant using algorithms involving DL primarily through CNN, hence its result will emerge after the classification.

4) Post-processing & Reporting:

The system will provide the results to the user, who may be a medical professional or technician, with relevant information like tumor characteristics and further actions to be taken.

B. System Architecture

The system architecture is created to accommodate the large volumes of ultrasound images. These images will be passed through deep learning models, thus achieving an efficient and accurate classification of the tumours. • Data Collection:

The ultrasound images are uploaded via the user interface

1) Preprocessing:

Image resizing, noise reduction, and enhancement are carried out as preprocessing steps to ensure the images are at their best for classification.

2) Model Inference:

The images that are preprocessed are sent to a trained DL model, for an instance if we consider, CNN, for classification. The model is able to detect patterns and features indicative of cancerous cells.

3) Result Generation:

The model generates a classification result, which is benign or malignant, and further details such as type, and stage.

4) Post-Processing:

The system provides the final report, which can be used by healthcare professionals for diagnosis and treatment planning.

C. Use Case Analysis

Figure 1 shows the functionalities performed by an user in the proposed system. User: An actor who actually uploads the image of the ultrasound scan, initiates the prediction module, and prints out the generated output. 1) Uploading Image: • User description:
A user in the system that uploads the images of a particular patient's ultrasound scan. Image formatpng or jpg are used during the uploading

1) Purpose:

Processes the ultrasound in the image which can be taken further for a computation.

2) Description:

This is where the user uploads the image and further processing starts. The signal will be received by the machine learning model which will process the image and evaluate.

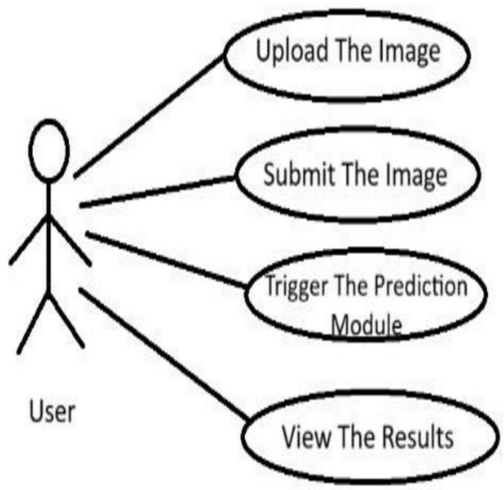


Fig. 1. Use Case Diagram

D. Trigger Prediction Module:

1) Description:

As soon as the image gets uploaded, then the prediction module will get activated. The system uses CNN for classifying whether the ultrasound image of the breast has a cancerous or non-cancerous tumor.

2) Purpose:

This will be seen as the main functionality of the system and a deep learning model predicts that the tumor of the breast is either cancerous or not.

E. View Results

1) Description:

This is the point of time after prediction, where, as a matter of fact, the user will view what the outcome it comprises and whether it falls under benign or malignant classification, and its approximated stage; the system will also give a description on the recommendation concerning the location of the tumor, any information relating to the stage, or any detail describing the tumor.

2) Purpose:

Here, the class result by the user forms the output which would help in any decision that could be made either to continue the treatment or discontinue the same.

F. Flow Chart

The breast cancer detection is presented in Figure 2, depicting various steps to be followed. Starting from data acquisition to the prediction step, it begins with the download of a dataset containing labeled ultrasound images that serve as the training basis for CNN model. This includes resizing and normalizing images, together with denoising the images into proper formats accepted by the model.

After these pre- processing operations, the sets of data used will have to be split further into one taken as training while another is being considered for the validation process and performance evaluation purposes of the model. Figure 2 shows the flow chart/flow followed in the proposed work.

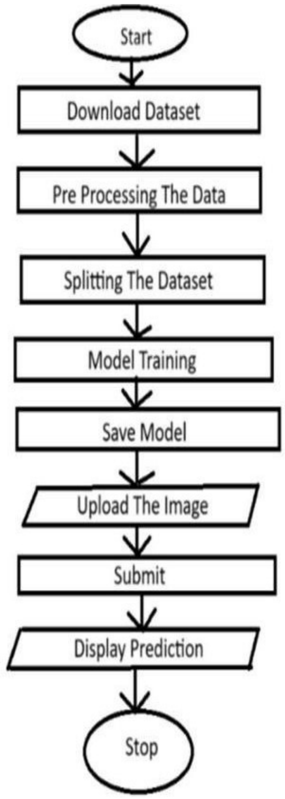


Fig. 2. Flow Chart for proposed system

Training of a model feeds the CNN architecture with a training dataset, hence learning about the design and characteristics. The trained model is then saved to use in the subsequent processes so as not to be trained again. Once the model is ready, end users can work with this system by using the interface where they upload an ultrasound image. The system processes an input, uses the trained model in order to predict the result, and then presents the outcome- determining whether the tumor is benign or malignant. The last output is the display of results that marks the conclusion of the process. In doing so, this way it guarantees an accurate and timely prediction and aiding the decision-making of clinicians.

IV. RESULTS AND DISCUSSION

The model is trained for 10 epochs, and several key performance metrics were monitored to analysis the correctness of the model. A detailed breakdown of training and validation accuracy, loss, and other evaluation metrics observed during and after training are as follows.

1) Training and Validation Accuracy: Training Accuracy was trending upwards throughout the epochs. In the first epoch, the model had an accuracy of 60.58%. This is expected because the model starts learning from the data. As training progressed, the accuracy continued to improve gradually until the tenth epoch, reaching a value of 83.40%. A high increase in accuracy indicated that the model correctly learned the training data, and its predictions improved with every iteration of the training process. Figure 3, highlights the training process involved.

   Validation accuracy also improved, apparently from 76.44% in the initial epoch to 85.89% by the tenth epoch. This shows that the model was performing well for trained images and generalizing well for unseen data as well. The drastic improvement in validation accuracy act as an major indicator of how the model is being able to distinguish between the features that actually required to indicate the tumour as cancerous or non-cancerous.
2) Loss Metrics: At Epoch 1 the training loss was 0.7608 and it went on decreasing consistently and at the end of Epoch 10 it reached up to 0.3880.

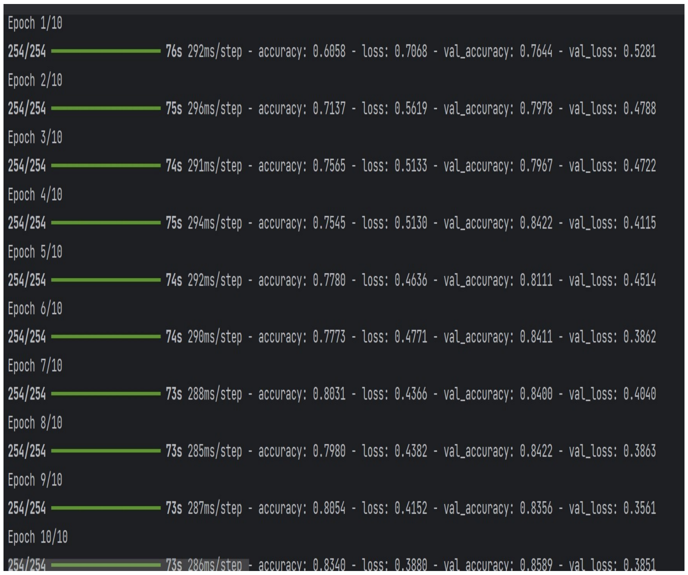


Fig.3. Trainingthemodel

Fig.4. ModelEvaluationMetricsonValidationSet

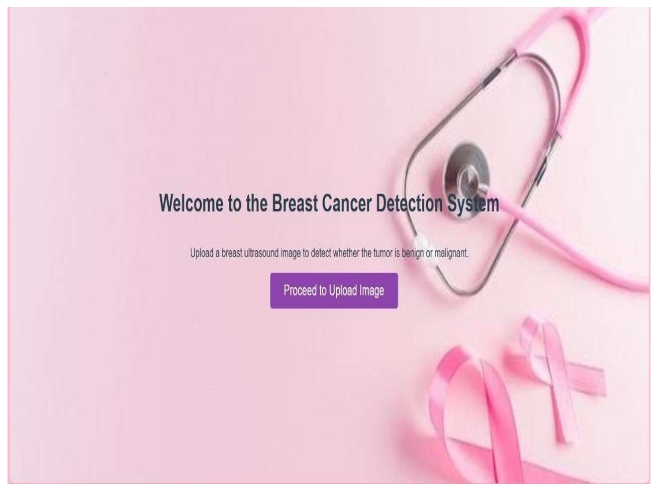


Fig. 5. Home Page

This consistent decrease confirms that the process of optimizing the model has been successfully as the model was able to understand the pattern of the trained data throughout its learning process. These smaller values of training loss indicate that the model is performing effectively in order to minimize the error between actual prediction and the prediction made by the system. Thus, giving more accurate results.

Similarly the validation loss also decreased. It started with 0.5281 in Epoch 1 and in the course of training process validation loss dropped till 0.3851 at the end of Epoch 10.This decrease in validation loss is very important as it indicates that system not only fits to trained data but it is only able to process the unseen data effectively and accurately.

Interpretation: In order to sum this up, at the end of Epoch 10 a consistent decrease has been observed both in training loss and validation loss. An significant increase has been noted in the training accuracy which went on from 76.44% in Epoch 1 to 85.89% till the end of Epoch 10. Not only in training accuracy, similar rise has been observed in validation accuracy also which was initially 60.58% and grew up to 83.40%. Figure 4, highlights the process of validation in the proposed approach. Figure 5, shows the homepage details for the proposed approach. Figure 6, depicts the sanpshot, for detection of cancer (stage 2) for one of the input patient record using the proposed model.

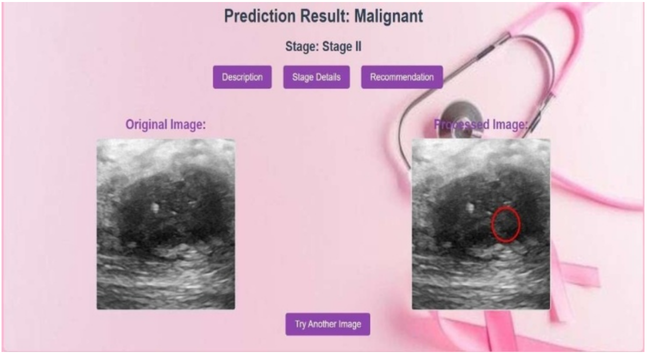


Fig. 6. Cancer Detection - Stage 2

A. Testing

The testcases that we have conducted checks the following cases that includes, firstly test has been conducted to check whether a valid benign is been able to predict or not and the system was successful in this case. Next we tested the system working for valid malignant prediction, the system was passed in this test by making correct malignant prediction. Further the system was tested for valid file format being uploaded or not ,if invalid file format is uploaded it prompts a message for user. Next testcase involved tests whether the tumour cells have been highlighted or if the prediction is malignant. So for this test case when an malignant ultra sound image was uploaded, the system successfully highlighted the tumour region. The set of test cases involed in testing the model is tabulated in Table I.

TABLE I Test Cases

Test Case ID	Input	Expected Output	Actual O/P	Remarks
1	Upload benign ultrasound image	”Benign” prediction, no processed image	Pass	Validates benign detection
2	Validates benign detection	Malignant” stage prediction	Pass	Ensures correct malignant prediction and staging
3	Upload nonimage file	File rejection, Prompt for valid image	Pass	Tests file format validation
4	Submit without file	Prompt to upload an image	Pass	Ensures form validation when no file is uploaded
5	Upload Large Image File	Image is resized to the required dimensions (128 X 128)	Pass	Tests whe ther the application handles larg e image files efficiently.

Further test cases involved testing whether the system can resize the image and make prediction or not. So when the system was tested for above test case by uploading an image of larger size, it was able to resize it and predict the result.

V. CONCLUSION

This work involved building a deep-learning model to detect breast cancer using ultrasound pictures. CNN architecture has been used in order to train the model using the ultrasound images so that it could be able differentiate between the tumour being benign or malignant. The model has been successfully trained and processed with an accuracy of 85.89% which indicates that the model can predict the result with minimum errors.

In order to make this system more user friendly a web application has been developed using python web framework called Flask. Using this application the user can easily get the prediction regarding whether the tumour in the breast is cancerous or non-cancerous via the uploaded ultrasound images.

In addition to predictions, the model also provides some recommendation for the classifications made. In case of malignant tumor it highlights the tumour region on the uploaded image and displays it as an processed image which serves as an excellent feedback for medical professionals and doctors, in order to examine and treat the patients in a more effective way.

Therefore, this work can be considered as a great means of tool for doctors for detecting breast cancer in the early stages itself. This model has been working well for the tests done so far. Further this model can be tested with more sophisticated and complex data in order to improve its reliability and making it more robust & useable for real time application.

REFERENCES

[1] Smita Raut, Shraddha Patil, Gopichand Shelke, “Automatic Detection of Lung Cancer Using Machine Learning(ML) Algorithms and Digital Image Processing.” Journal of Biomedical Engineering and Analysis, Vol. 6, Issue 1, January 2021, ISSN (Online) 2456-0774.

[2] Vikul J. Pawar, Kailash D. Kharat, Suraj R. Pardeshi, Prashant D. Pathak “Advancements in Image Processing Techniques for Lung Cancer Detection.” Journal of Diagnostics and Medical Imaging, Vol. 9, No.4, July – August 2020, ISSN 2278-3091.

[3] Eali Stephen, Neal Joshua, Midhun Chakravarthy, Debnath Bhattacharyya. “Evaluating Accuracy Levels of Machine learning(ML) Algorithms for Lung Cancer Nodule Classification.” Journal of Computational Oncology, Vol. 18, Issue 2, July 2023, pp. 85-98. ISSN: 5678-1234. Vol. 34, No. 3, June, 2020, pp. 351359, May 2020.

[4] M. Tahmooresi, A. Afshar, B. Bashari Rad, K.B. Nowshath, M.A. Bamiah ”A Hybrid Model for Effective Breast Cancer Detection Using Machine Learning(ML) Algorithms” Journal of Health Informatics and Medical Imaging, ISSN: 2289-8131, Vol. 10, 2024.

[5] Komal Sharma, Shruti Gujral, Akwinder Kaur “Automated Tumour Detection in Brain MRI Using Machine Learning Algorithms” Journal of Diagnostic Applications and Medical Imaging, Volume 103 – No.1, October 2014

[6] Rasool Fakoor, Faisal Ladhak Unsupervised Feature Learning for Cancer Detection and Classification Using Gene Expression Data.” Journal of Computational Biology and Medical Diagnostics, Vol. 19, Issue 2, February 2025, pp. 75-88. ISSN: 1234-5678.

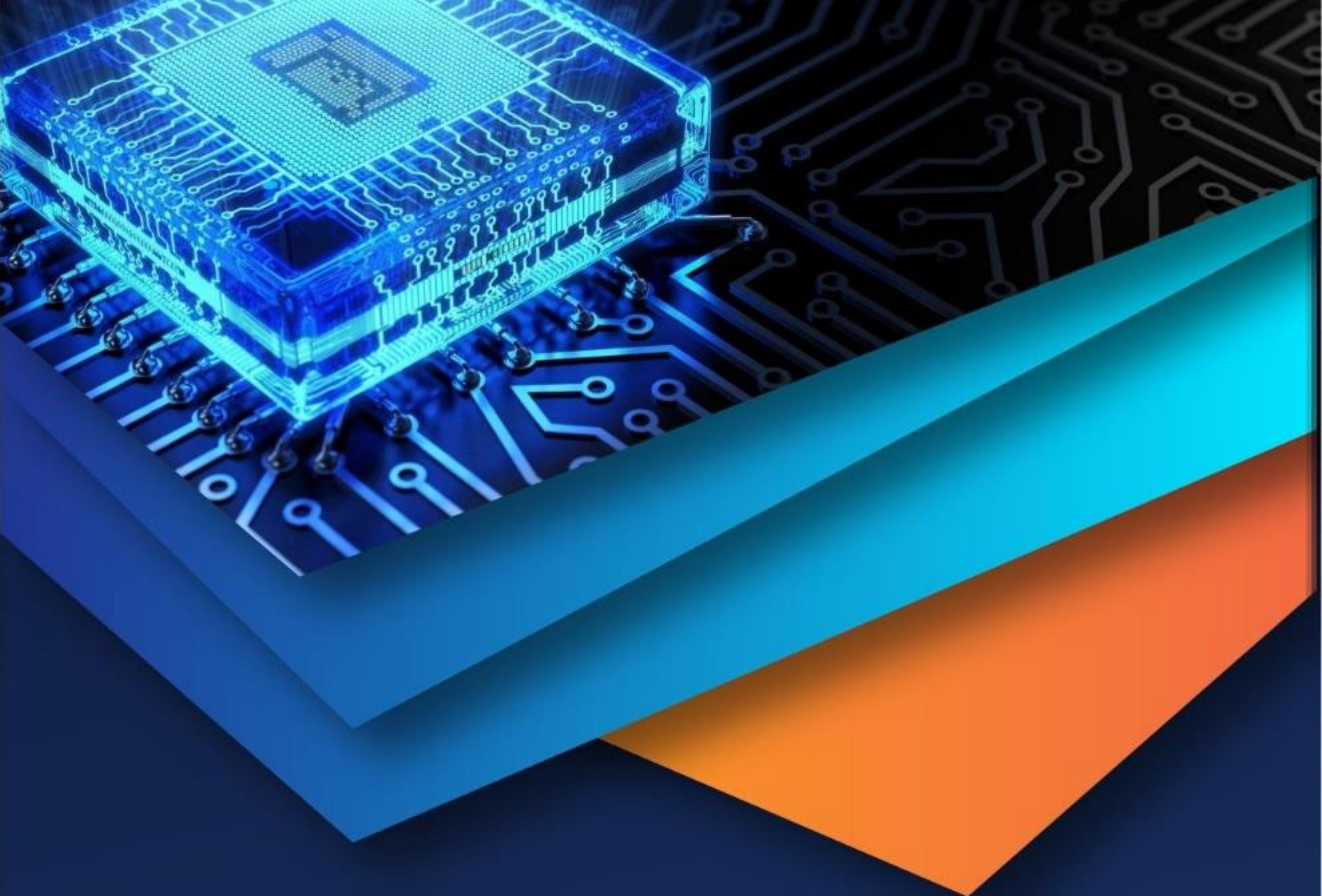
[7] Shrutika Gadekar, Sweta Bhise, Simran Bepari, Deepmala kale, Aishwarya Singh Gour, Dr. Shailendra Aswale, “Automated Diagnosis of Breast Cancer Using Machine Learning Techniques.” Journal of Cancer Detection and Predictive Analytics, ISSN: 2278-0181 Vol. 10 Issue 07, July-2021

[8] Aaswad Sawant, Mayur Bhandari, Ravikumar Yadav, Rohan Yele, Sneha Bendale Brain, “Cancer Detection Using MRI and TensorFlow-Based Convolutional Neural Networks.” Journal of Artificial Intelligence and Medical Imaging, Vol. 05 Issue: 04, Apr-2018, e-ISSN: 2395-0056, p-ISSN: 2395-007, February 2025

[9] Javaria Amin, Muhammad Sharif, Mudassar Raza, Tanzila Saba, Muhammad Almas Anjum “Early Detection of Brain

[10] Tumor Using Fused Features from LBP and Gabor Wavelet Transform.” Journal of Biomedical Engineering and Analysis, Vol. 6, Issue 1, January 2021, ISSN 2456-0774.

[11] Jinen Daghrir, Lotfi Tlig, Moez Bouchouicha, Mounir Sayadi “Hybrid Method for Melanoma Skin Cancer Detection Using Machine Learning and Data Fusion” Journal of Skin Cancer Research and Advanced Diagnostics, Vol. 24, Issue 1, March 2025, pp. 40-55. ISSN: 6789-1234.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY