iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VIII **Month of publication:** Aug 2023

DOI: https://doi.org/10.22214/ijraset.2023.55503

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Breast Cancer Detection and Prediction using Image Processing and ML

Dwarakanath G V[1], Megha Bhairagond[2]
[1]Associate Professor, [2]Student, Department of Master of Computer Applications BMS Institute of Technology and Management Bengaluru, Karnataka, India

Abstract: Breast cancer is one of the most common and potentially life-threatening forms of cancer that affects a significant number of women worldwide. Early detection of breast cancer plays a crucial role in improving patient outcomes and survival rates. Machine learning algorithms have the potential to analyze large volumes of medical imaging data, extract meaningful features, and assist in the identification of suspicious regions or potential tumors. By leveraging these algorithms, healthcare professionals can make more accurate and timely diagnoses, leading to improved patient care and outcomes. Breast cancer is a prevalent form of cancer that affects a significant number of women worldwide. Early detection plays a crucial role in improving patient outcomes and survival rates. Breast cancer detection refers to the process of identifying abnormal changes in breast tissue, such as tumors or growths, that may indicate the presence of cancer cells. Early detection plays a vital role in improving patient outcomes by allowing for timely intervention and targeted treatment plans. Several techniques are utilized in breast cancer detection, including screening mammography, clinical breast examination, and breast self-examination. Mammography, the most common method, involves using low-dose X-rays to capture images of the breast tissue. It can detect tumors or suspicious areas even before they can be felt by a physician or the patient.

In recent years, advancements in medical imaging and machine learning techniques have shown promising results in breast cancer detection. The designing of the model began with classification of Histopathological image dataset into Cancerous and Non - cancerous classes using Support Vector Machine (SVM) and Convolutional Neural Network (CNN) algorithms. Both the classifiers are examined on the basis of sensitivity, specificity..

Keywords: Breast Cancer detection, images, Convolutional Neural Networks.

I. INTRODUCTION

Cancer detection has always been a severe challenge in diagnosis and treatment planning for pathologists and healthcare professionals. Manual cancer detection using microscopic biopsy images is subjective; results may differ from expert to expert based on their experience and other factors such as the lack of specific and reliable quantitative metrics to classify biopsy images as normal or cancerous. When biologically interpretable and clinically meaningful feature-based techniques are applied for disease detection, the automated identification of malignant tissue from microscopic biopsy pictures helps to alleviate the difficulties mentioned above and provides improved outcomes. Many models have been developed and deployed in recent years, each with benefits and drawbacks. Histopathologists examine the particular traits in the cells and tissue architecture to discover and diagnose cancer from microscopic biopsy photos. Shape and size of cells, shape and size of cell nuclei, and cell distribution are frequent parameters used for cancer detection and diagnosis from microscopic biopsy images.

II. RELATED WORK

The sources related to the cancer cells and related images. They cover different methodologies, evaluation techniques, and challenges that serve as a foundation for this research paper on cancer detection. This section provides an overview on the work related to Breast Cancer Detection and Prediction.

1) Automated Detection and Classification of Breast Cancer Tumor Cells using Machine Learning and Deep Learning on Histopathological Images :This paper they considered Histopathological data set and performed classification using SVM and CNN models and achieved accuracy as 99% for train test ratio 60:40 to design the framework. For segmentation GA and K-Mean is applied and observed that performance of GA is better than K-means. If the training dataset is not representative of the broader population or lacks certain subtypes of breast cancer, the models may struggle to generalize well to new, unseen cases. In this section, the analysis of result is covered and comparison of result with other existing classification algorithms.

The input data set is histopathological images and they are given to SVM and CNN classification models to classify the image whether it is benign or malignant. To compare the result of both the algorithm quantitative analysis parameters is considered that are discussed in the preliminary section. To analyses the result we have considered three different train test ratio 70:30, 60:40, and 80:20. Further if the image is having cancerous cell then segmentation is applied to the image for the extraction of cancerous cell. For segmentation K-Mean and GA is used and their results are compared as well.

2) Automated malignancy detection in breast Histopathological images: This paper the results presented in Table 4 summarize our experimental trials. Given the 27-dimensional set of features the classification accuracy combining textural features, network features and morphometric features reaches 86.43 ± 4.73. We note that individual features achieve an inferior best classification accuracy: 80.43±3.16 for the feature encoding the average shortest path between nuclei and 81.00 ± 6.53 for the textons based on H channel from pre-processed images. Combining these texton features, network features and features encoding nuclei statistics proves an advantage over individual method classification. This is because the complementary information provided by the different individual features captures different cancer manifestations in breast histopathology The Maximum Relevance - Minimum Redundancy (MR-MR) method used for feature selection, proves it's usefulness. Adding more features to the data set used for classification saturates the classification accuracy since the features are not exhibiting any more orthogonal properties. Analyzing which of the 8 features taken into consideration for the MR-MR feature selection, reveals a combination of network features, text on features and morphometric features.
3) Breast cancer classification using deep belief networks: This paper In this research, they presented an automatic diagnosis system for detecting breast cancer based on DBN unsupervised pre-training phase followed by a supervised back propagation neural network phase (DBNNN). The pre-trained back propagation neural network with unsupervised phase DBN achieves higher classification accuracy in comparison to a classifier with just one supervised phase. The rationale behind this enhancement could be that the learning of input statistics from input feature space by DBN phase initializes back propagation neural network to search objective function near a good local optima in supervised learning phase. From this experiment at the specified network architecture, DBN-NN complex accuracy outperforms RIW-BPNN when back propagation neural network uses conjugate gradient algorithm for learning. DBN-NN still outperforms RIW-BPNN when we use Levenberg-Marquardt for training in back propagation neural network phase. The enhancement of overall neural network accuracy is reaching 99.68% with 100% sensitivity and 99.47% specificity in breast cancer case.
4) Mammogram classification using dynamic time warping: This paper presents a new approach for breast cancer classification using time series analysis. In particular, the region of interest (ROI) in mammogram images is classified as normal or abnormal using dynamic time warping (DTW) as a similarity measure. According to the analogous case in time series analysis, the DTW subsumes Euclidean distance (ED) as a specific case with increased robustness due to DTW flexibility to address local horizontal vertical deformations. This method is especially attractive for biomedical image analysis and is applied to mammogram classification for the first time in this paper. The current study concludes that varying the size of the ROI images and the restriction on the search criteria for the warping path do not affect the performance because the method produces good classification results with reduced computational complexity. The method is tested on the IRMA and MIAS dataset using the k-nearest neighbour classifier for different k values, which produces an area under curve (AUC) value of 0.9713 for one of the best scenarios.
5) An enhanced breast cancer diagnosis scheme based on two-step-SVM technique: This paper proposes an automatic diagnostic method for breast tumour disease using hybrid Support Vector Machine (SVM) and the Two-Step Clustering Technique. The hybrid technique is aimed at improving the diagnostic accuracy and reducing diagnostic miss-classification, thereby solving the classification problems related to Breast Tumour. To distinguish the hidden patterns of the malignant and benign tumours, the Two-Step algorithm and SVM have been combined and employed to differentiate the incoming tumours. The developed hybrid method enhances the accuracy by 99.1% when examined on the UCI-WBC data set. Moreover, in terms of evaluation measures, it has been shown experimentally results that the hybrid method outperforms the modern classification techniques for breast cancer diagnosis.
6) Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis: In this paper, we have provided explanations of different ML approaches and their applications in BC diagnosis and prognosis used to analyse the data in the benchmark database WBCD. ML techniques have shown their remarkable ability to improve classification and prediction accuracy. Various methods have been shown in Table 2 with references, algorithms, sampling strategies and classification accuracies, providing a clear and intuitive catalogue of information. Although lots of algorithms have achieved very high accuracy in WBCD, the development of improved algorithms is still necessary.

Classification accuracy is a very important assessment criteria but it is not the only one. Different algorithms consider different aspects, and have different mechanisms. Although for several decades ANNs have dominated BC diagnosis and prognosis, it is clear that more recently alternative ML methods have been applied to intelligent healthcare systems to provide a variety of options to physicians.

7) Breast Cancer Prediction Using Machine Learning: In this paper they are going to make use of four algorithms and compare the results. The algorithms are k- Nearest Neighbour (k-NN), Support Vector Machines (SVM), Random Forest and Naive Bayes Classifier. The dataset that we are going to use for our research is labelled which means that all the rows of data are being organized and have a distinct column name. According to the nature of thier dataset, they will carry out Supervised Learning. In Supervised Learning, algorithms take a known set of input and output data and train machine learning models to predict the new output data based on unseen input data.
8) Breast cancer detection using artificial intelligence techniques: In this paper they arranged it in ascending chronological order. This work found that ANNs were first used in the field of HIA around 2012. ANNs and PNNs were the most frequently applied algorithms. However, in feature extraction, most of the work used textural and morphological features. It was clear that Deep CNNs were quite effective for early detection and diagnosis of breast cancer, leading to more successful treatment. Prediction of Non-Communicable Diseases (NCDs) was conducted using many algorithms. In , the authors compared the performance of various classification algorithms. The classification algorithms were performed on eight NCD datasets using eight classification algorithms and a 10-fold cross-validation method. These were evaluated using AUC as an indicator of accuracy. The authors stated that the NCD datasets have noisy data and irrelevant attributes. KNN, SVM and NN proved to be robust to this noise. In addition, they stated that the irrelevant attribute problem can be handled with some pre-processing techniques to improve the accuracy rate.
9) Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms:In this paper the study is based on genetic programming and machine learning algorithms that aim to construct a system to accurately differentiate between benign and malignant breast tumors. The aim of this study was to optimize the learning algorithm. In this context, we applied the genetic programming technique to select the best features and perfect parameter values of the machine learning classifiers. The performance of the proposed method was based on sensitivity, specificity, precision, accuracy, and the roc curves. The present study proves that genetic programming can automatically find the best model by combining feature preprocessing methods and classifier algorithms.
10) Comparative Analysis of Various Machine Learning Techniques for Diagnosis of Breast Cancer: In this paper they compared the classification results obtained from the techniques i.e. KNN, SVM, Random Forest, Decision Tree (Recursive Partitioning and Conditional Inference Tree). the cancer forms in either the lobules or the ducts of the breast. Cancer also can occur within the adipose tissue or the fibrous connective tissue within your breast. The uncontrolled cancer.The dataset used was Wisconsin Breast Cancer dataset obtained from UCI repository. Simulation results showed that KNN was the best classifier followed by SVM, Random Forest and Decision Tree.
11) Breast Cancer Prediction using Deep learning and Machine Learning Techniques: In this paper presented a novel method to detect breast cancer by employing techniques of Machine Learning that is Logistic Regression, Random Forest, K-Nearest Neighbor, Decision tree, Support Vector Machine and Naïve Bayes Classifier and techniques of Deep Learning that is Artificial Neural Network, Convolutional Neural Network and Recurrent Neural Network. To separate the two classes of data points, there are many possible hyper planes that could be chosen. The objective is to find a plane that has the maximum margin The comparative analysis between the Machine Learning and Deep learning techniques concluded that the accuracy obtained in the case of CNN model (97.3%) and ANN model (99.3%) was more efficient than the Machine Learning models.
12) A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications: In this paper they used data visualization and machine learning techniques including logistic regression, k-nearest neighbors, support vector machine, naïve Bayes, decision tree, random forest, and rotation forest were applied to this dataset. R, Minitab, and Python were chosen to be applied to these machine learning techniques and visualization. A comparative analysis was performed amongst the all the techniques. Results obtained with the logistic regression model with all features included showed the highest classification accuracy (98.1%), and the proposed approach revealed the enhancement in accuracy performances
13) Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis: In this paper they conducted a performance comparison between different machine learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naïve Bayes (NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) dataset.

Experimental results showed that SVM gives the highest accuracy (97.13%) with lowest error rate. SVM has been applied to breast cancer diagnosis, where it learns to distinguish between malignant and benign tumors based on features extracted from medical images or other data. All experiments are executed within a simulation environment and conducted in WEKA data mining tool.

14) Breast Cancer Detection by Leveraging Machine Learning: In this paper they presented a novel method to detect breast cancer by employing techniques of Machine Learning such as Naïve Bayes classifier, SVM classifier, Bi-clustering Ada Boost techniques, RCNN classifier and Bidirectional Recurrent Neural Networks (HA-BiRNN). A comparative analysis was done between the Machine learning techniques and the proposed methodology (Deep Neural Network with Support Value) and the simulated results concluded that the DNN algorithm was advantageous in both performance, efficiency and quality of images are crucial in the latest medical systems whilst the other techniques didn't perform as expected.

III. METHODOLOGY.

This paper proposes an experiment using cancer images takenfrom the Kaggle platform. The dataset comprises 4k RGB images with an 80x80 pixel resolution, categorized into two classes: “Malignant" and "Benign." Various image preprocessing techniques were subsequently applied to improve the image quality. Afterward, the Convolutional Neural Network (CNN) model was trained using the sample images. The figure below illustrates the proposed breast cancer detection model.

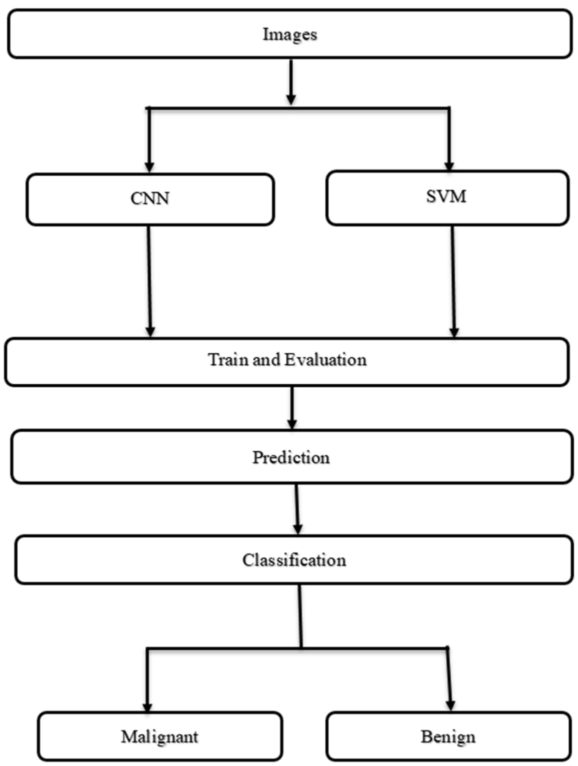


FIGURE 1: Proposed Model for breast cancer detection and Prediction.

A. Dataset

The dataset used in this project is microscopic biopsy images of breast cancer[Fig.1] and microscopic biopsy images of non-cancerous cells[Fig.2] derived from the Breast Cancer Histopathological Database (BreakHis). The Breast Cancer Histopathological Image Classification (BreakHis) is made up of 1693 microscopic images of breast tumor tissue collected from 82 patients, 547 of which are benign and 1147 of which are malignant (700X460 pixels, 3- 8-bit depth in each channel, PNG format).

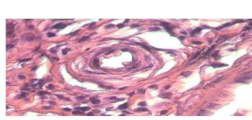
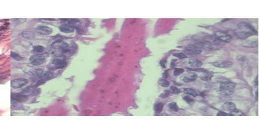



Fig.1: Cancerous tissue Fig.2: Non-Cancerous tissue

1) Image Preprocessing and Model Architecture

To comprehend image processing, one must first learn the concept of an image. A picture can be represented using the 2D function F(x,y), where x and y are spatial coordinates. The amplitude of F at a specific value of x,y equals the intensity of a picture at that place. The x, y, and amplitude values in a digital image are all finite. It's a pixel array divided into rows and columns. Pixels are image elements that store color and intensity information. The spatial coordinates x, y, and z can also be used to depict a picture in 3D Pixels are layered in a matrix format. An RGB image is what this is called.

There are various types of images - RGB images and Grayscale images.

Image processing is the application of operations to an image to improve it or derive helpful information from it. It is a form of signal processing in which a picture serves as the input, and the output is either that image or its characteristics/features. Image processing is one of today's most rapidly changing technologies. It's also an important area of research in engineering and computer science.

The three phases that makeup image processing are as follows:

- Importing an image using image acquisition tools.
- Analyzing and altering the image.
- Producing an output that can be an altered image or a report based on image analysis.

2) Convolutional Neural Network

A CNN or Convolutional Neural Network is a deep learning neural network designed to analyze structured arrays of data-like representations. When one thinks about neural networks, one usually thinks of matrix multiplications, but this isn't the case with ConvNet. It employs a technique known as Convolution. Convolution is a mathematical operation on two functions that yields a third function that explains how the shape of one is changed by the other.

CNN's are excellent at detecting unique features in input images, such as lines, gradient circles, and even eyes and faces. Because of this feature, convolutional neural networks are effective in computer vision. CNN does not require pre-processing and can run straight on an under-done image.

Feedforward neural network with up to 20 layers is known as a Convolutional Neural Network. Convolutional Neural Network strength stems from a layer known as the convolutional layer.

3) Python Libraries and Modules

- TensorFlow: TensorFlow is an open-source deep learning framework developed and maintained by the Google Brain team. It is one of the most popular and widely used libraries for building, training, and deploying machine learning and deep learning models.
- Keras: Keras is a popular open-source deep learning framework written in Python. It was developed and maintained by François Chollet, initially as an independent project, but later integrated into the TensorFlow ecosystem. Keras provides a user-friendly and high-level API for building, training, and deploying deep learning models, making it accessible to both beginners and experienced researchers or engineers.
- CV2: cv2 (OpenCV) stands for Open Source Computer Vision Library. It is a powerful open-source library that provides a wide range of tools and functions for computer vision tasks and image processing. OpenCV is written in C++ and has a Python interface, making it accessible to developers working in both languages.
- Scipy: SciPy is an open-source scientific computing library for Python that builds on top of NumPy (Numerical Python). It provides a wide range of algorithms and functions for scientific and engineering computations, signal processing, optimization, statistics, and more. SciPy is an essential tool for researchers, engineers, and data scientists working on various scientific and technical projects.
- GC: Garbage Collection is an automated memory management technique used by programming languages and runtime environments to reclaim memory occupied by objects that are no longer in use and cannot be accessed by the application. The main goal of garbage collection is to prevent memory leaks and ensure efficient memory utilization.
- Itertools: is a module in Python's standard library that provides a set of fast, memory-efficient tools for working with iterators. It offers various functions to create, combine, and manipulate iterators in a concise and efficient manner. Itertools is part of the Python standard library.

- Tqdm: is a popular Python library that provides a fast and extensible progress bar for loops and iterable objects. The name "tqdm" stands for "taqaddum," which is an Arabic word meaning "progress" or "progression."
  The library is particularly useful when you have tasks that take a significant amount of time to complete, and you want to monitor their progress in real-time.
- PIL: stands for Python Imaging Library, was an open-source library for image processing in Python. However, as of my last update in September 2021, the original PIL library is no longer actively developed and has been replaced by the Pillow library.
- Functools: is a Python standard library module that provides higher-order functions and operations on functions. It contains useful tools for working with functions and is especially handy when dealing with functional programming concepts. The module is part of the Python standard library, so no additional installation is required to use it.
- Collections: is a built-in Python standard library module that provides alternative data structures and additional data manipulation utilities compared to the built-in Python data structures like lists, tuples, and dictionaries. It contains various container datatypes and utility functions to handle them efficiently. The collections module is designed to be memory-efficient and optimized for specific use cases

B. Model Training

A separate dataset known as test dataset is prepared to cross verify the accuracy of trained model, which contains images that the model has never seen during training. There are some common evaluation methods to determine the accuracy of the model, They are as follows:

a) Accuracy: The accuracy metric represents the correct predictions made by the model over the total number of test samples. It is the most straightforward evaluation metric and provides an overall view of the performance of the model. A = (n) / (N) Where A represents Accuracy, n and N signifies the number of correct predictions and the total amount of test samples respectively.

b) Confusion Matrix: A confusion matrix is a table that shows the amount of correct and incorrect predictions detected in the model for each class (ship detection). It provides more detailed information about the model's performance and helps identify which ships are frequently misidentified.

c) ROC Curve: The Receiver Operating Characteristic (ROC) is a graphical representation of the model's performance at various classification thresholds. It plots the true positive rate (TPR or recall) against the false positive rate (FPR) as the threshold for classifying a sample is varied.

d) Precision: Precision is the ratio of true positive predictions to the total amount of positive predictions made by the model. It measures how many predicted positive samples were actually positive. Precision = (True Positives) / (True Positives + False Positives)

C. User Interface

The User Interface (UI) module focuses on providing an interactive platform for users to experience . The UI module is responsible for providing user friendly interface and also provides good command over the system.

D. Model Testing

A separate dataset known as test dataset is prepared to cross verify the accuracy of trained model, which contains images that the model has never seen during training. There are some common evaluation methods to determine the accuracy of the model, They are as follows:

a) Accuracy: The accuracy metric represents the correct predictions made by the model over the total number of test samples. It is the most straightforward evaluation metric and provides an overall view of the performance of the model. A = (n) / (N) Where A represents Accuracy, n and N signifies the number of correct predictions and the total amount of test samples respectively.

b) Confusion Matrix: A confusion matrix is a table that shows the amount of correct and incorrect predictions detected in the model for each class (ship detection). It provides more detailed information about the model's performance and helps identify which ships are frequently misidentified.

c) ROC Curve: The Receiver Operating Characteristic (ROC) is a graphical representation of the model's performance at various classification thresholds. It plots the true positive rate (TPR or recall) against the false positive rate (FPR) as the threshold for classifying a sample is varied.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

E. Ship Detection

Once this model is trained and validated, it can be deployed in real-time applications. In such applications, the ship module captures images from the device, preprocesses the images, and passes them through the trained model to highlight the ships in that area.

F. User Interface

The User Interface (UI) module focuses on providing an interactive platform for users to experience . The UI module is responsible for providing user friendly interface and also provides good command over the system.

IV. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

CNN and SVM has been implemented on the set of 1693 microscopic images of breast tumor tissue collected from 82 patients, 547 of which are benign and 1147 of which are malignant breast cancer tissue images using DenseNet 201 framework using Python v3.6. 2 models have been employed to determine the best accuracy score in order to detect the cancerous and non-cancerous tissues. Model 1 has an accuracy of 93%, whereas Model 2 has an accuracy of 96%.security threats, and respond swiftly to illegal activities, smuggling, or piracy.

To evaluate the effectiveness of the proposed system comprehensive tests were conducted. The efficiency of the system was assessed with the help of various metrics.

1) Accuracy: The accuracy is typically measured as the percentage of correctly classified facial expressions out of the total number of expressions in the dataset or during real-time testing. For example, if the model correctly identifies 80 out of 100 images with ships , the accuracy would be 80%.

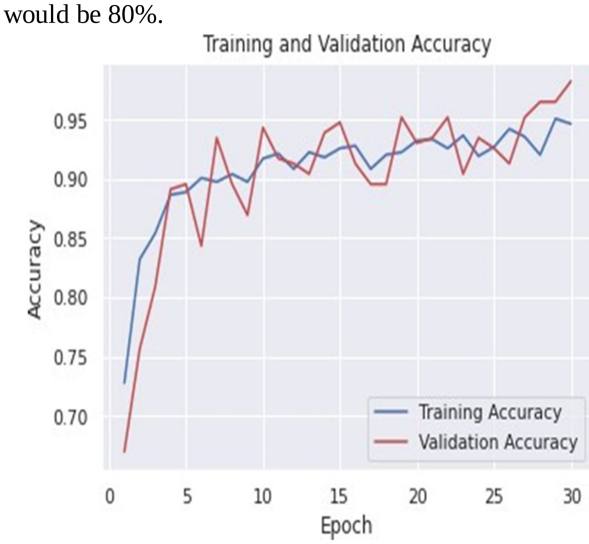


FIGURE 3: accuracy of the model

2) Confusion Matrix: A confusion matrix is a table that shows the number of correct and incorrect predictions made by the model for each class. It provides more detailed information about the model's performance and helps identify which facial expressions are frequently misclassified.

Following are the components of the confusion matrix.

a) True Positive (TP): The amount of samples that belong to a particular class and are correctly predicted as that class by the model.
b) False Positive (FP): The amount of samples that do not belong to a particular class, but are incorrectly predicted as that class by the model.
c) True Negative (TN): The amount of samples that do not belong to a particular class and are correctly predicted as not belonging to that class by the model.
d) False Negative (FN): The amount of samples that belong to a particular class, but are incorrectly predicted as not belonging to that class by the model.

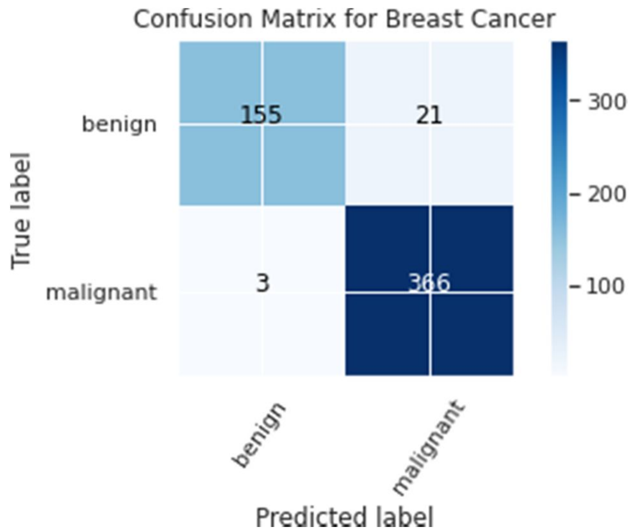


FIGURE 4: Representation of Confusion Matrix

3) Precision: Precision is calculated as the fraction of true positive predictions (correctly predicted positive samples) to the total amount of positive predictions made by the model (true positive plus false positive). It represents how well the model avoids false positives and correctly identifies positive samples for a particular class.

Mathematically, precision is defined as:

Precision = True Positives / (True Positives + False Positives)

V. FINDINGS AND IMPLICATIONS OF THE RESEARCH

The findings and implications of research on breast cancer detection using image processing and machine learning are significant and have the potential to revolutionize the way breast cancer is diagnosed and treated.

Improved Accuracy: Research has shown that combining image processing techniques, such as segmentation and feature extraction, with machine learning algorithms can lead to higher accuracy in detecting breast cancer from medical images like mammograms and ultrasound scans.

Early Detection: Machine learning models trained on large datasets of breast images can identify subtle patterns and anomalies that might not be easily discernible to human radiologists. This can lead to earlier detection of breast cancer, which is crucial for improving patient outcomes.

Reduced False Positives/Negatives: By fine-tuning algorithms and leveraging deep learning approaches, researchers have been able to reduce the rate of false positives and false negatives in breast cancer detection, thus minimizing unnecessary biopsies or missed diagnoses.

Personalized Medicine: Machine learning algorithms can analyze a patient's medical history, genetic information, and imaging data to tailor treatment plans specifically to each individual. This can optimize the efficacy of treatments while minimizing side effects.

Automation and Efficiency: Implementing machine learning in breast cancer detection can significantly speed up the analysis process. This is particularly important in regions where there's a shortage of radiologists, allowing for faster diagnosis and treatment initiation.

Integration with Clinical Workflow: Integrating machine learning tools into the existing clinical workflow can enhance the decision-making process for healthcare providers. These tools can provide additional insights and recommendations to aid radiologists in making accurate diagnoses.

Implications:

Enhanced Diagnosis: The research has the potential to enhance the accuracy and reliability of breast cancer diagnosis, leading to improved patient outcomes and survival rates.

Resource Optimization: By reducing the need for extensive manual analysis, healthcare resources can be optimized, leading to reduced costs and quicker turnaround times for diagnoses.

Global Impact: This research can have a significant impact in regions where access to skilled radiologists is limited. Automated or assisted diagnosis through machine learning can extend quality healthcare services to underserved populations.

Research Focus: The findings indicate the importance of further research in refining existing algorithms, exploring new imaging techniques, and incorporating more diverse and comprehensive datasets to continually improve the accuracy and applicability of the models.

Ethical Considerations: As with any medical technology, there are ethical considerations surrounding data privacy, bias in algorithms, and the responsible implementation of AI in clinical settings. Addressing these concerns is crucial for the responsible deployment of these tools.

In conclusion, the research findings in breast cancer detection using image processing and machine learning have shown immense promise in improving diagnostic accuracy, early detection, and personalized treatment. While there are still challenges to overcome, the implications are far-reaching and have the potential to positively impact the field of oncology and patient care.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

It's a difficult task to automate breast cancer screening to improve patient care. Above, SVM and CNN architectures were compared for the detection of breast cancer from the dataset of 1693 microscopic images of breast tumor tissue collected from 82 patients, 547 of which are benign and 1147 of which are malignant (700 X 460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format). The proposed system, which employs Model 2, has a 96% accuracy rate in comparison to Model 1, with an accuracy of 93%. The main scope of this project is for healthcare and oncologists to diagnose cancer accurately as early as possible and to reduce human mistakes in the diagnosis phase. In the future, the usage of AI and ML for efficient diagnosis should be implied to decrease human errors and help people fight cancer as early as possible

B. Future Work

Multi-Modal Data Fusion: Incorporating additional imaging modalities, such as ultrasound or MRI, could enhance the accuracy and reliability of breast cancer detection. Fusion of data from multiple sources might provide a more comprehensive view of the tissue under examination.

Explainability and Interpretability: Developing methods to interpret and explain the decisions made by the machine learning models will be crucial, especially in a clinical setting. Techniques like attention maps and feature visualization can help doctors understand the basis for the model's predictions.

Large-Scale Clinical Validation: Conducting extensive clinical trials and validation studies on diverse patient populations is essential before deploying the developed system in real-world healthcare scenarios. This will ensure the reliability and generalizability of the proposed approach.

Incremental Learning and Continual Training: Implementing mechanisms for continual learning and model adaptation can ensure that the system remains up-to-date with evolving medical knowledge and new imaging data.

Data Augmentation and Transfer Learning: Exploring advanced data augmentation techniques and transfer learning strategies can improve the model's performance, especially when dealing with limited annotated data.

Privacy and Security Considerations: Addressing patient data privacy concerns and implementing robust security measures will be critical in the deployment of such systems to protect sensitive medical information.

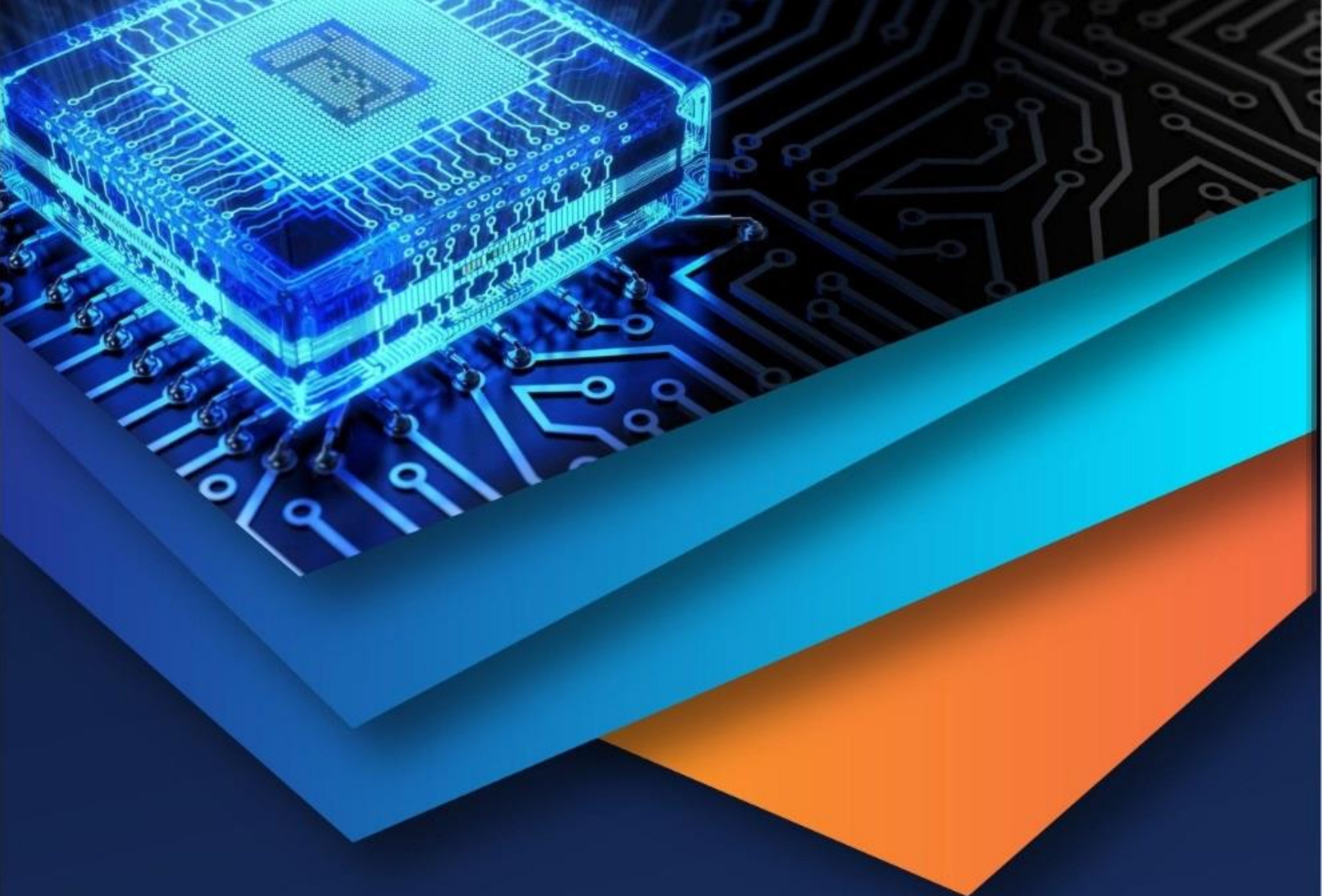
Real-Time Detection and Decision Support: Developing real-time detection systems that can provide immediate decision support to radiologists during image interpretation can significantly expedite the diagnostic process.

Integration into Clinical Workflow: Seamless integration of the developed system into the existing clinical workflow, including picture archiving and communication systems (PACS) and electronic health records (EHR), will facilitate its adoption by healthcare professionals.

REFERENCES

[1] Ramik Rawal "Breast Cancer Prediction Using MachineLearning"(2020).
[2] S. Karthik, R. Srinivasa Perumal and P. V. S. S. R. Chandra Mouli "Breast Cancer Classification Using Deep Neural Networks"(2019).
[3] Abdullah-Al Nahid, Aaron Mikaelian and Yinan Kong"Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation." (2018).
[4] Syed Jamal Safdar Gardezi, Ahmed Elazab, Baiying Lei and Tianfu Wang "Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review" (2019).
[5] Sebastien Jean Mambou , Petra Maresova , Ondrej Krejcar , Ali Selamat and Kamil Kuca.
[6] "Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model" (2018).

[7] Hannah Le "Using Machine learning models for breast cancer detection" (2018).
[8] Saleem Z. Ramadan “Methods Used in Computer- Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review” (2020).
[9] M. Tahmooresi , A. Afshar, B. Bashari Rad , K. B. Nowshath and M. A. Bamiah "Early Detection of Breast Cancer Using Machine Learning Techniques".
[10] Mandeep Rana, Pooja Chandorkar, Alishiba Dsouza and Nikahat Kazi “Breast Cancer Diagnosis and Recurrence Prediction Using Machine Learning Techniques” (2015).
[11] Shubham Sharma , Archit Aggarwal and Tanupriya Choudhury "Breast Cancer Detection Using Machine Learning Algorithms" (2018).
[12] Ram MurtiRawat ,Shivam Panchal ,Vivek Kumar Singh ,Yash Panchal "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning" (2020)



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY