

**DEEP LEARNING APPROACH FOR**

**BREAST CANCER CLASSIFICATION:**

**A NAAN MUDHALVAN REPORT**

**SUBMITTED BY:**

**SURYA R**

**(711721244057)**

**IN PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE OF**

**BACHELOR OF TECHNOLOGY**

**IN**

**DEPARTMENT OF COMPUTER SCIENCE AND BUSINESS SYSTEM**

**KGISL INSTITUTE OF TECHNOLOGY**

**ANNA UNIVERSITY: CHENNAI 600025**

**MAY 2024**



**ANNA UNIVERSITY: CHENNAI 600025**

**MAY 2024**

**BONAFIDE CERTIFICATE**

Certified that this Naan Mudhalvan report "**Deep Learning Approach For Breast Cancer Classification**" is the bonafide work of "**SURYA R**" who belongs to III Year B.Tech Computer Science and Business System during the Sixth Semester of Academic Year 2023-2024.

**FACULTY IN CHARGE HEAD OF THE DEPARTMENT**

Certified that the candidates were examined by us for Naan Mudhalvan Practical Viva held on\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ at KGiSL Institute of Technology, Saravanampatti, Coimbatore 641035.

**INTERNAL EXTERNAL EXAMINER EXTERNAL**

**ACKNOWLEDGEMENT**

We express our deepest gratitude to our **Chairman** and **Managing Director**, **Dr. Ashok Bakthavachalam**, for providing us with an environment to complete our Internship project successfully.

We are grateful to our **CEO of Academic Initiatives**, **Mr. Aravind Kumar Rajendran**, and our beloved **Secretary**, **Dr. Rajkumar N**. Our sincere thanks to honorable **Principal**, **Dr. Suresh Kumar**, and **Academic Director**, **Dr. Shankar P**, for their support, guidance, and blessings.

We would like to thank **Dr. Rajkumar P**, **Head of the Department**, and **Ms. Ranjithaa R**, **Naan Mudhalvan Coordinator**, Department of Computer Science and Business System, for firm support during the entire course of this Internship and for modeling us both technically and morally for achieving greater success in this project work.

We also thank all the faculty members of our department for their help in making this Internship project a successful one. Finally, we take this opportunity to extend our deep appreciation to our Family and Friends, for all they meant to us during the crucial times of the completion of our project.

**TABLE OF CONTENT:**

|  |  |
| --- | --- |
| 1. | ABSTRACT |
| 2. | INTRODUCTION |
| 3. | DOMAIN |
| 4. | METHODOLOGY |
| 5. | SOLUTION AND VALUE PROPOSITION |
| 6. | KEY FEATURES |
| 7. | MODELLING APPROACH |
| 8. | SOURCE CODE |
| 9. | CONCLUSION |

**ABSTRACT:**

Our project outlines a deep learning-based solution for breast cancer classification, highlighting its multifaceted benefits. By leveraging advanced algorithms, the solution promises heightened diagnostic accuracy, streamlining the process and leading to more precise outcomes. Through automation, it enhances efficiency, reducing the time required for diagnosis and enabling prompt treatment initiation. Additionally, the solution offers significant cost savings by minimizing manual review and interpretation of images, optimizing resource allocation within healthcare institutions. Notably, it surpasses human performance in specific diagnostic aspects, ensuring superior reliability and effectiveness. Furthermore, its scalability and customization make it adaptable to various healthcare settings, from small clinics to large hospitals, catering to diverse needs. In essence, this solution represents a transformative advancement in breast cancer diagnosis, empowering healthcare professionals and ultimately improving patient care and outcomes.

**INTRODUCTION:**

Breast cancer remains a daunting global health issue despite medical advancements. The reliance on manual interpretation of mammography and histopathology images introduces subjectivity and variability into diagnosis. Thus, there's an urgent need for automated systems to enhance diagnostic accuracy and efficiency. Our project aims to address this challenge by developing a deep learning-based system for automated breast cancer classification. Leveraging convolutional neural networks (CNNs), we seek to analyze digital mammography and histopathology images to classify tumors as benign or malignant. Through data collection, preprocessing, model development, training, and validation phases, our goal is to create a tool that assists healthcare professionals in making more accurate and timely diagnoses. By standardizing diagnostic processes and minimizing errors, our solution holds the potential to significantly improve patient outcomes. In this context, the development of automated systems represents a critical advancement in combating breast cancer, offering hope for better care and outcomes globally.

**DOMAIN:**

The domain of breast cancer diagnosis and treatment encompasses a multidisciplinary approach involving medical professionals, researchers, and technology developers. It involves the understanding, detection, diagnosis, and treatment of breast cancer, one of the most prevalent forms of cancer affecting women globally. This domain integrates various fields such as oncology, radiology, pathology, genetics, and computational biology.

In recent years, there has been a growing emphasis on leveraging technological advancements, particularly in artificial intelligence (AI) and machine learning, to improve the accuracy and efficiency of breast cancer diagnosis. Deep learning-based solutions, utilizing techniques like convolutional neural networks (CNNs), have shown promise in automating the analysis of mammography and histopathology images, thereby aiding in early detection and classification of tumors.

Furthermore, the domain of breast cancer diagnosis and treatment extends beyond medical interventions to encompass aspects of patient care, public health initiatives, and advocacy for research funding. Collaborative efforts between healthcare providers, researchers, policymakers, and patient advocacy groups are essential for advancing knowledge, improving outcomes, and ultimately reducing the burden of breast cancer on individuals and society.

**METHODOLOGY:**

The methodology employed in this project involves a deep learning-based approach for breast cancer diagnosis, integrating various stages from data collection to model evaluation.

1. **Data Collection:** Relevant datasets containing digital mammography and histopathology images are gathered from sources like medical institutions or publicly available repositories.
2. **Data Preprocessing:** The collected data undergoes preprocessing steps such as image resizing, normalization, and augmentation to enhance quality and consistency.
3. **Model Development:** A deep learning model is developed using convolutional neural networks (CNNs), chosen for their effectiveness in image classification tasks. The architecture is designed to analyze mammography and histopathology images and classify tumors as benign or malignant.
4. **Training and Validation:** The model is trained on a portion of the dataset and validated on a separate subset to optimize its performance and prevent overfitting.

5. **Evaluation:** The trained model is evaluated using various metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in breast cancer classification.

**SOLUTHON AND VALUES PROPOSED:**

The solution proposed is a deep learning-based system for breast cancer diagnosis, offering a range of benefits to healthcare professionals, patients, and healthcare institutions.

1. **Improved Diagnostic Accuracy:** By leveraging advanced algorithms, the solution enhances the accuracy of breast cancer diagnosis by analyzing digital images with high precision, detecting subtle patterns indicative of cancerous lesions.
2. **Enhanced Efficiency:** Automation of image analysis reduces the time required for diagnosis, enabling prompt initiation of treatment and improving patient outcomes. This efficiency gains significance in the context of breast cancer, where early detection is crucial for successful treatment.
3. **Cost Savings:** By minimizing the need for manual review and interpretation of images, the solution optimizes resource utilization within healthcare institutions, leading to significant cost savings. These savings can be allocated to other critical areas of patient care and research.
4. **Empowering Healthcare Professionals:** The user-friendly interface and intuitive design of the solution empower healthcare professionals, such as radiologists and pathologists, to integrate the tool seamlessly into their existing workflow.

**KEY FEATURES:**

1. **Advanced Diagnostic Accuracy:** The solution utilizes state-of-the-art deep learning algorithms to analyze digital mammography and histopathology images with remarkable precision, detecting subtle patterns indicative of breast cancer.
2. **Efficient Automation:** By automating the image analysis process, the solution reduces the time required for diagnosis, enabling healthcare professionals to make timely decisions regarding patient care and treatment initiation.
3. **Cost-Efficiency:** Automation minimizes the need for manual review and interpretation of images, leading to significant cost savings for healthcare institutions by optimizing resource utilization and reducing operating expenses.
4. **User-Friendly Interface:** The solution features a user-friendly interface with an intuitive design, making it easy for healthcare professionals, including radiologists and pathologists, to integrate the tool seamlessly into their existing workflow without the need for extensive training in deep learning techniques.
5. **Scalability and Customization:** The solution is scalable and can be customized to suit the specific needs and preferences of different healthcare settings, ranging from small clinics to large hospitals, ensuring adaptability and flexibility in implementation.

**MODELLING APPROACH:**

The modeling approach for the deep learning-based solution involves several key steps to develop an effective system for breast cancer diagnosis:

1. **Data Collection:** Relevant datasets containing digital mammography and histopathology images are gathered from medical institutions or publicly available repositories.
2. **Data Preprocessing:** The collected data undergoes preprocessing steps such as image resizing, normalization, and augmentation to enhance quality and consistency, ensuring optimal performance during model training.
3. **Model Architecture:** A deep learning model is designed using convolutional neural networks (CNNs), chosen for their effectiveness in image classification tasks. The architecture is tailored to analyze mammography and histopathology images and classify tumors as benign or malignant.
4. **Training and Validation:** The model is trained on a portion of the dataset and validated on a separate subset to optimize its performance and prevent overfitting. This involves adjusting parameters and hyperparameters to achieve the desired level of accuracy and generalization.

**Input Layer:**

Receives data with 30 features.

**Flatten Layer:**

Converts input data into a 1D array (output

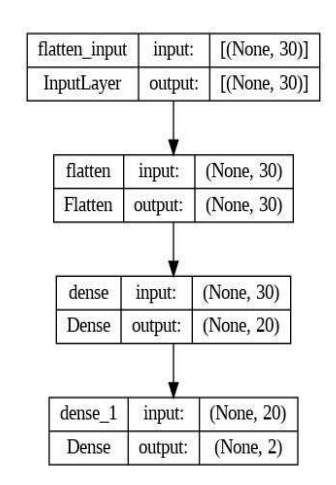
shape: (None, 30)).

**Dense Layers:**

Two dense layers:

First layer reduces dimensionality to 20.

Second layer produces final output (shape: (None, 2)).



**SOURCE CODE:**

**Importing Dependencies:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import sklearn.datasets

from sklearn.model\_selection import train\_test\_split

**Data Collection and Processing:**

breast\_cancer\_dataset = sklearn.datasets.load\_breast\_cancer()

print(breast\_cancer\_dataset)

**Loading Data To Dataframe:**

data\_frame= pd.DataFrame(breast\_cancer\_dataset.data, columns = breast\_cancer\_dataset.feature\_names)

data\_frame.head()

**Adding Target Column To Dataframe:**

data\_frame["label"] = breast\_cancer\_dataset.target

data\_frame.tail()

data\_frame.shape

**Getting Information Of Dataset:**

data\_frame.info()

**Checking Null Values:**

data\_frame.isnull().sum()

**Statistical Measures Of Dataset:**

data\_frame.describe()

**Distribution Of Data:**

data\_frame["label"].value\_counts()

data\_frame.groupby("label").mean()

**Separating Features And Target:**

X = data\_frame.drop(columns = "label", axis = 1)

Y = data\_frame["label"]

print(X)

print(Y)

#### Data Standardization:

#### from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

standardized\_data = scaler.fit\_transform(X)

print(standardized\_data)

**Split Train and Test data:**

print("Total data: ", X.shape)

print("Train data: ", X\_train.shape)

print("Test data: ", X\_test.shape)

**Importing Tensorflow and Keras:**

# setting up the layers of Neural Network

model = keras.Sequential([

keras.layers.Flatten(input\_shape = (30,)),

keras.layers.Dense(20, activation = 'relu'),

keras.layers.Dense(2, activation = "sigmoid")

])

# compiling the Neural Network

model.compile(optimizer = "adam",

loss = "sparse\_categorical\_crossentropy",

metrics = ["accuracy"])

# training the Neural Network

history = model.fit(X\_train, Y\_train, validation\_split = 0.1, epochs = 10)

# visualize train validation loss

plt.plot(history.history["loss"])

plt.plot(history.history["val\_loss"])

plt.title("Model Accuracy")

plt.ylabel("loss")

plt.xlabel("epoch")

plt.legend(["training", "validation"], loc = "upper right")

loss, accuracy = model.evaluate(X\_test, Y\_test)

print(accuracy)

print(X\_test.shape)

print(X\_test[0])

Y\_pred = model.predict(X\_test)

**Predictive System:**

sample = data\_frame.drop("label", axis = 1).iloc[0].values

input\_data = (sample)

print(input\_data)

# reshape the array as we are predicting for one instance

input\_data\_reshaped = input\_data.reshape(1, -1)

# standardize the input data as model is trained as such

std\_data = scaler.transform(input\_data\_reshaped)

print(std\_data)

prediction = model.predict(std\_data)

print(prediction)

predictin\_label = [np.argmax(prediction)]

print(predictin\_label)

if(predictin\_label[0] == 0):

print("The tumor is Malignant.")

else:

print("The tumor is Benign.")

input\_data = X\_test[0]

print(input\_data)

# reshape the array as we are predicting for one instance

input\_data\_reshaped = input\_data.reshape(1, -1)

print(input\_data\_reshaped)

prediction = model.predict(input\_data\_reshaped)

print(prediction)

predictin\_label = [np.argmax(prediction)]

print(predictin\_label)

if(predictin\_label[0] == 0):

print("The tumor is Malignant.")

else:

print("The tumor is Benign.")

for i in range(1,5):

input\_data = X\_test[i]

print(input\_data)

# reshape the array as we are predicting for one instance

input\_data\_reshaped = input\_data.reshape(1, -1)

print(input\_data\_reshaped)

prediction = model.predict(input\_data\_reshaped)

print(prediction)

predictin\_label = [np.argmax(prediction)]

print(predictin\_label)

if(predictin\_label[0] == 0):

print("The tumor is Malignant.")

else:

print("The tumor is Benign.")

**DATASET AND OUTPUT:**

[ 0.23650097 -0.04413248 0.20846223 0.09217991 -0.45832546 -0.11581219

-0.36910756 -0.01880758 0.24600557 -0.60639571 -0.42398132 -0.48779257

-0.34407559 -0.34030218 -0.56968273 -0.24189769 -0.54820881 0.01198128

-0.64441597 -0.37064995 -0.00397387 0.08350252 0.05476982 -0.12443056

-0.04684277 0.31002241 -0.44035741 0.52217753 0.09907982 0.04290197]

[[ 0.23650097 -0.04413248 0.20846223 0.09217991 -0.45832546 -0.11581219

-0.36910756 -0.01880758 0.24600557 -0.60639571 -0.42398132 -0.48779257

-0.34407559 -0.34030218 -0.56968273 -0.24189769 -0.54820881 0.01198128

-0.64441597 -0.37064995 -0.00397387 0.08350252 0.05476982 -0.12443056

-0.04684277 0.31002241 -0.44035741 0.52217753 0.09907982 0.04290197]]

1/1 [==============================] - 0s 41ms/step

[[0.5984049 0.516898 ]]

[0]

The tumor is Malignant.

[-1.26650216 -0.27683919 -1.27314927 -1.04830523 -0.94296212 -0.94872971

-0.92918482 -1.1130141 -0.48053217 0.18746257 -0.2294449 0.1726286

-0.29259241 -0.42588883 1.31306027 -0.72142993 -0.71565178 -1.27531463

0.77341331 -0.30861673 -1.06009478 -0.17216223 -1.07679418 -0.87962986

0.28192641 -0.81912559 -0.99835737 -1.36439788 0.25276898 -0.28793046]

[[-1.26650216 -0.27683919 -1.27314927 -1.04830523 -0.94296212 -0.94872971

-0.92918482 -1.1130141 -0.48053217 0.18746257 -0.2294449 0.1726286

-0.29259241 -0.42588883 1.31306027 -0.72142993 -0.71565178 -1.27531463

0.77341331 -0.30861673 -1.06009478 -0.17216223 -1.07679418 -0.87962986

0.28192641 -0.81912559 -0.99835737 -1.36439788 0.25276898 -0.28793046]]

1/1 [==============================] - 0s 42ms/step

[[0.4043982 0.3581323]]

[0]

The tumor is Malignant.

[ 2.31264474 0.08851035 2.5056397 2.42917903 2.5790156 3.26835349

4.23856683 3.44042339 2.71769425 1.07630031 2.92283913 0.59537073

3.55627059 2.85280594 -0.17533341 3.42054383 2.3908297 2.10935226

1.28073224 0.46982453 2.51207887 0.37987822 2.96484624 2.60068572

1.65398981 2.83358867 3.30362995 2.68587702 1.86731403 0.77272829]

[[ 2.31264474 0.08851035 2.5056397 2.42917903 2.5790156 3.26835349

4.23856683 3.44042339 2.71769425 1.07630031 2.92283913 0.59537073

3.55627059 2.85280594 -0.17533341 3.42054383 2.3908297 2.10935226

1.28073224 0.46982453 2.51207887 0.37987822 2.96484624 2.60068572

1.65398981 2.83358867 3.30362995 2.68587702 1.86731403 0.77272829]]

1/1 [==============================] - 0s 59ms/step

[[0.9922254 0.31641027]]

[0]

The tumor is Malignant.

[ 0.0831333 -0.63986166 0.08983445 -0.03893322 0.08182025 0.18115749

-0.10959631 -0.17253971 0.3664867 0.18746257 -0.20237583 -0.80965718

-0.0935901 -0.18034975 -0.8213596 0.28234619 0.215729 0.10281869

-0.21943042 -0.1595857 0.08300079 -0.67860643 0.12327725 -0.03249173

-0.13013097 0.52694605 0.49563943 0.40797805 0.24629786 0.20471616]

[[ 0.0831333 -0.63986166 0.08983445 -0.03893322 0.08182025 0.18115749

-0.10959631 -0.17253971 0.3664867 0.18746257 -0.20237583 -0.80965718

-0.0935901 -0.18034975 -0.8213596 0.28234619 0.215729 0.10281869

-0.21943042 -0.1595857 0.08300079 -0.67860643 0.12327725 -0.03249173

-0.13013097 0.52694605 0.49563943 0.40797805 0.24629786 0.20471616]]

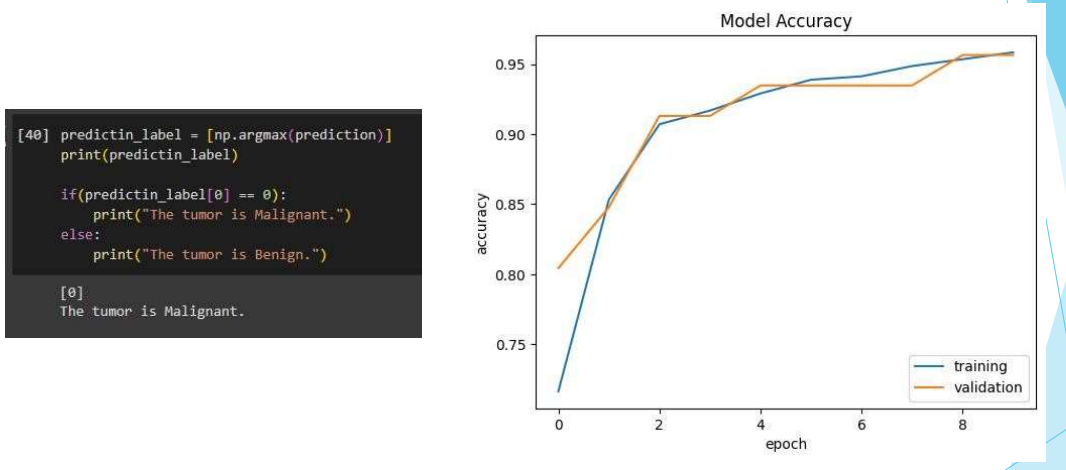
1/1 [==============================] - 0s 34ms/step

[[0.69207525 0.48300433]]

[0]

T

The tumor is Malignant.



**CONCLUSION:**

In conclusion, the development of a deep learning-based solution for breast cancer diagnosis represents a significant advancement in healthcare technology. Through the integration of advanced algorithms and machine learning techniques, the solution offers a promising avenue for improving diagnostic accuracy, efficiency, and patient outcomes.

By automating the analysis of digital mammography and histopathology images, the solution streamlines the diagnostic process, enabling healthcare professionals to make more informed decisions in a timely manner. The ability to surpass human performance in certain aspects of diagnosis and the customization options for different healthcare settings further enhance its utility and value proposition.

Moreover, the cost savings achieved through automation and the empowerment of healthcare professionals underscore the broader impact of this solution on healthcare delivery and resource utilization. Overall, the deep learning-based solution for breast cancer diagnosis holds immense potential to revolutionize clinical practice, contributing to improved patient care, reduced healthcare costs, and ultimately, better outcomes for individuals affected by this devastating disease.