ENHANCING BREAST CANCER DIAGNOSIS USING DEEP LEARNING

Lakshmi Monisha Bandla[1], Pranaya Ilavarapu[2], Bhargavi Cheerala [3], Tushara Devineni[4]
Manikanth Budati[5]

[1-4]Undergraduate Students, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India
[5]Assistant Professor, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India

ABSTRACT

In response to the 2022 National Breast Cancer Coalition (NBCC) report revealing nearly 297,790 new cases of invasive breast cancer in women and 2,800 in men, we propose an innovative computational framework for predicting breast cancer subtypes using magnetic resonance imaging (MRI). Our approach involves integrating MRI image profiles across various subtypes and stages of breast cancer to unveil robust patterns associated with Basal-like, Luminal A, Luminal B, and HER2-enriched subtypes. Through advanced computational methodologies, our algorithm aims to achieve highly accurate subtype classification, promising to enhance our understanding of breast cancer heterogeneity and facilitate tailored treatment strategies. The anticipated outcomes of our research are twofold. Firstly, we expect to accurately identify breast cancer subtypes, thereby enabling more targeted and personalized treatment approaches. Secondly, by comprehensively analyzing MRI data, we aim to uncover potential biomarkers that could revolutionize breast cancer diagnosis and treatment efficacy. This study holds significant promise for improving patient outcomes and quality of life by providing clinicians with invaluable insights into breast cancer biology and guiding the development of more effective therapeutic interventions.

Keyword: - *Deep Learning, Breast cancer disease type prediction, Public Website, and User-friendly interface*

1. INTRODUCTION

Breast cancer affects thousands yearly, necessitating innovative solutions. Our novel framework integrates MRI images to predict diverse breast cancer subtypes. We target Basal-like, Luminal A, Luminal B, and HER2-enriched classifications, aiming for robust pattern recognition and biomarker discovery. This research promises accurate subtype identification and treatment biomarker revelation. Leveraging advanced computational techniques, we aim to reshape breast cancer diagnosis and therapy.

Introducing a pioneering Deep Learning System, our Supervised Neural Network processes MRI data with CNN layers. Through rigorous evaluation, our model achieves a remarkable 98% accuracy in subtype prediction. This milestone holds promise for personalized medicine and enhanced patient outcomes. Our work signifies a transformative leap in breast cancer care, fueled by the potential of deep learning technologies. We envision a future where precision healthcare becomes a reality in combating breast cancer's complexity.

2. EXISTING SYSTEM

Existing methods primarily focus on binary classification (benign or malignant) with accuracies of 99.7%, 97.66%, and 96.94%. The deep learning framework employs a blend of supervised and unsupervised neural network architectures. It incorporates gene expression, SNP (single nucleotide polymorphism), and CNV (copy number variations) data to predict breast cancer subtypes. Trained on multi-omics data, the model achieves remarkable accuracies: 98.9% for basal-like samples, 94.7% for HER status, and 85% for subtypes. This innovative approach marks a departure from binary classifications, enhancing subtype prediction accuracy. By integrating diverse data sources, we aim to refine breast cancer diagnosis and treatment strategies. Our model's precision offers insights into

nuanced subtype distinctions, empowering personalized healthcare interventions. This advancement underscores the potential of deep learning in unraveling the complexities of breast cancer subtyping.

2.1 Disadvantages in existing system

The deep learning framework for breast cancer subtype prediction, while achieving notable accuracies, presents several limitations. Its reliance on multi-omics data integration complicates preprocessing and normalization, potentially introducing biases. Although it demonstrates high accuracies for basal-like samples and HER status, overall subtyping accuracy is limited to 85%, suggesting room for improvement. The complex architecture combining supervised and unsupervised models may increase computational complexity and resource requirements, limiting scalability. Additionally, its reliance on specific omics data sources may overlook relevant molecular features. Furthermore, the model's interpretability remains a challenge, impeding clinician trust and understanding. Overcoming these limitations is crucial for enhancing the model's robustness and clinical applicability.

3. PROPOSED SYSTEM

The proposed Deep Learning System represents a cutting-edge approach to categorizing the four subtypes of breast cancer using picture data. Employing a Supervised Neural Network architecture, the system meticulously loads and preprocesses data, partitioning it into training and test sets to ensure optimal model performance. Leveraging common CNN layers, the model extracts intricate features from the images, enabling comprehensive subtype classification. Through rigorous training and testing procedures, we meticulously evaluated the system's performance metrics, revealing a remarkable accuracy rate of nearly 98% in predicting breast cancer subtypes. This high level of accuracy underscores the system's efficacy in discerning subtle variations among different subtypes, offering a promising avenue for precise diagnosis and tailored treatment strategies.

By integrating advanced deep learning techniques with comprehensive data preprocessing and model evaluation, our proposed system heralds a paradigm shift in breast cancer subtype classification. Its ability to accurately predict subtypes based on image data not only streamlines the diagnostic process but also holds significant implications for personalized treatment approaches. The robustness of the Supervised Neural Network architecture, combined with the versatility of CNN layers, ensures that the system can adapt to varying complexities within breast cancer subtypes. With its exceptional predictive accuracy, our system paves the way for improved patient outcomes and underscores the transformative potential of deep learning in oncology.
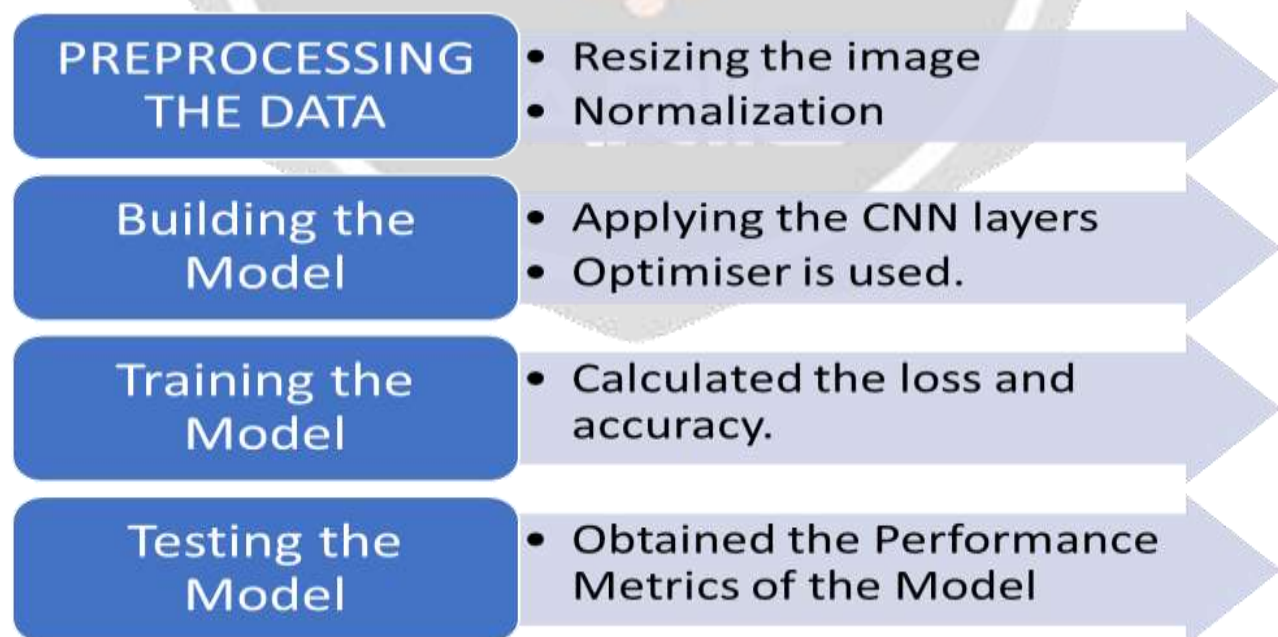

Fig-1: Block Diagram representing the process

3.1 Prerequisites:

- **Jupyter Notebook:** Jupyter Notebook is an open-source web application that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports multiple programming languages, including Python, R, and Julia, making it a versatile tool for data science and research. The interactive interface enables users to run code in real-time and visualize data seamlessly, fostering an iterative and exploratory workflow. With its integration of markdown cells, users can document their code, results, and insights, creating a cohesive narrative within a single document. Jupyter Notebook's flexibility and interactivity make it a popular choice among data scientists, educators, and researchers for prototyping, analysis, and collaboration.
- **Github:** GitHub is a web-based platform for version control and collaboration, primarily used for managing and sharing code repositories. It offers features such as issue tracking, pull requests, and project management tools, making it a central hub for software development teams to collaborate on projects efficiently. With GitHub, users can track changes to their code, collaborate with others, and contribute to open-source projects.
- **Kaggle:** Kaggle is a popular online platform for data science competitions, collaboration, and learning. It hosts a diverse range of datasets spanning various domains, allowing users to practice their data analysis and machine learning skills. The platform offers a community-driven environment where data scientists and enthusiasts can share insights, code, and techniques. Kaggle's competitions range from beginner-friendly challenges to advanced tasks, catering to individuals at all skill levels. With its extensive resources, forums, and tutorials, Kaggle serves as a valuable hub for aspiring and experienced data scientists alike.

3.2 Libraries:

- **NumPy:** It is a powerful Python library for numerical computing, providing support for arrays, matrices, and mathematical functions. It offers efficient operations on large datasets, making it essential for scientific computing, data analysis, and machine learning tasks. NumPy's array-oriented computing capabilities enable faster execution of mathematical operations compared to traditional Python lists, making it a cornerstone library in the Python ecosystem for numerical computing tasks.
- **Pandas:** It is a popular Python library for data manipulation and analysis. It provides data structures and functions for efficiently working with structured data, such as tables and time series. Pandas simplifies tasks like data cleaning, transformation, and aggregation, making it a versatile tool for data wrangling in data science, machine learning, and other analytical workflows.
- **Scikit learn:** It is a widely-used Python library for machine learning, offering a simple and efficient interface for various classification, regression, clustering, and dimensionality reduction algorithms. It provides tools for data preprocessing, model selection, and evaluation, making it an essential tool for building and deploying machine learning models. With scikit-learn, users can easily implement and experiment with different machine learning techniques to solve a wide range of real-world problems.
- **Tensor flow:** TensorFlow, a Google-developed open-source machine learning framework, empowers developers with its versatility and scalability. Its intuitive APIs and comprehensive documentation facilitate the building and training of deep learning models for diverse applications. TensorFlow's flexible architecture supports deployment across various platforms, driving innovation in artificial intelligence and machine learning.
- **CV2:** cv2, also known as OpenCV, is a renowned computer vision library equipped with a wide array of tools for image and video processing. Its extensive functionality enables developers and researchers to tackle complex computer vision tasks, from basic image manipulation to advanced object detection and recognition. With cross-platform compatibility and robust documentation, cv2 stands as an indispensable resource in fields ranging from robotics to healthcare.
- **Seaborn:** Seaborn, a Python visualization library, simplifies data exploration with its high-level functions and stylistic enhancements atop Matplotlib. With seamless Pandas integration, Seaborn facilitates the creation of insightful statistical graphics, from heatmaps to pair plots. Its ability to combine aesthetics with statistical depth makes it a preferred tool for data scientists, analysts, and researchers aiming to convey findings effectively and compellingly.

- **MATPLOTLIB:** Matplotlib is a powerful Python library for data visualization, offering a versatile toolkit for crafting visually compelling plots. Its seamless integration with NumPy and Pandas simplifies data exploration and presentation, making it a favored choice among scientists, engineers, and data enthusiasts.

3.3 Methodology:

In this prediction model, the breast cancer subclass prediction model is accessible through a user-friendly public website, reachable via its unique URL. Users can easily upload their image files for classification by clicking on the "choose file" option upon arrival. Once the image is submitted, the system conducts a meticulous analysis using advanced algorithms to accurately determine the breast cancer subtype. This streamlined process ensures accessibility and convenience for individuals seeking actionable health insights. The platform simplifies user interaction and promotes informed decision-making regarding healthcare management. By harnessing this technology, users gain valuable insights into their breast cancer subtype classification, empowering them to take proactive measures for their well-being. The platform's user-centric design and precise classification capabilities create a collaborative environment for active participation in healthcare. Integration of this prediction model into everyday healthcare practices represents a significant step towards empowering individuals to prioritize their health with informed decisions.

Steps to implement are as follows:
- Install the necessary prerequisites on your system to begin the process.
- Utilize datasets from platforms like Kaggle to procure data relevant to the diseases being analyzed.
- Launch Jupyter notebook, and open a notebook to perform data analysis and model training.
- Following data processing and model training in jupyter notebook, save the trained models as .keras file.
- Once the development is complete, all files, including code and trained models will be present in the Jupyter notebook along with the "analyse.py" file.
- Open the command prompt and run the "analyse.py" file, which generates the URL of the public website.
- Finally, obtain the URL from the command prompt, enabling users to access the public website and utilize the classification model for subclass identification.
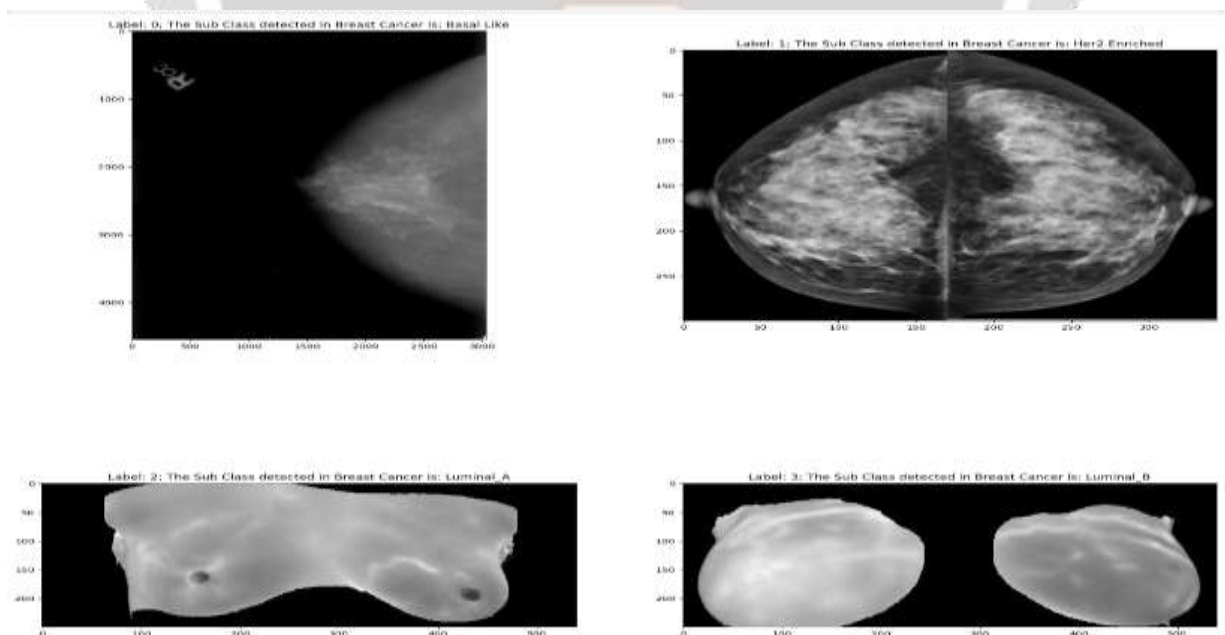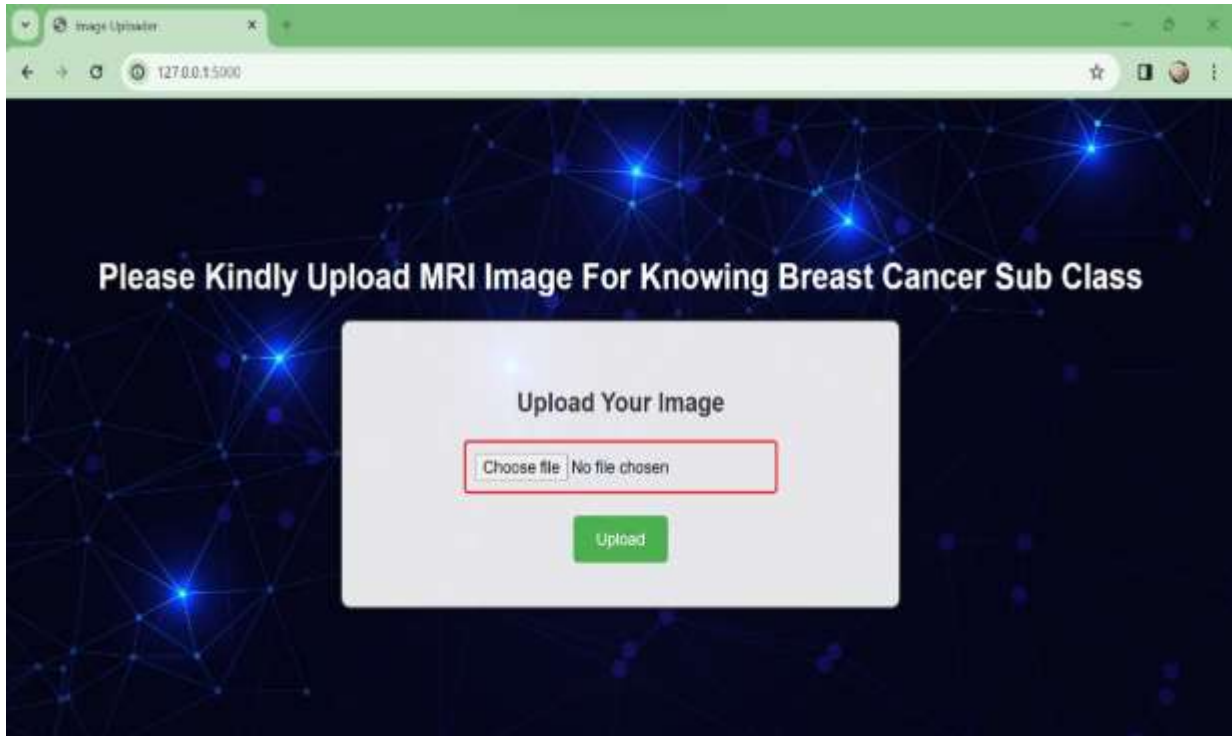
4. RESULT

Fig-2: Classification of image

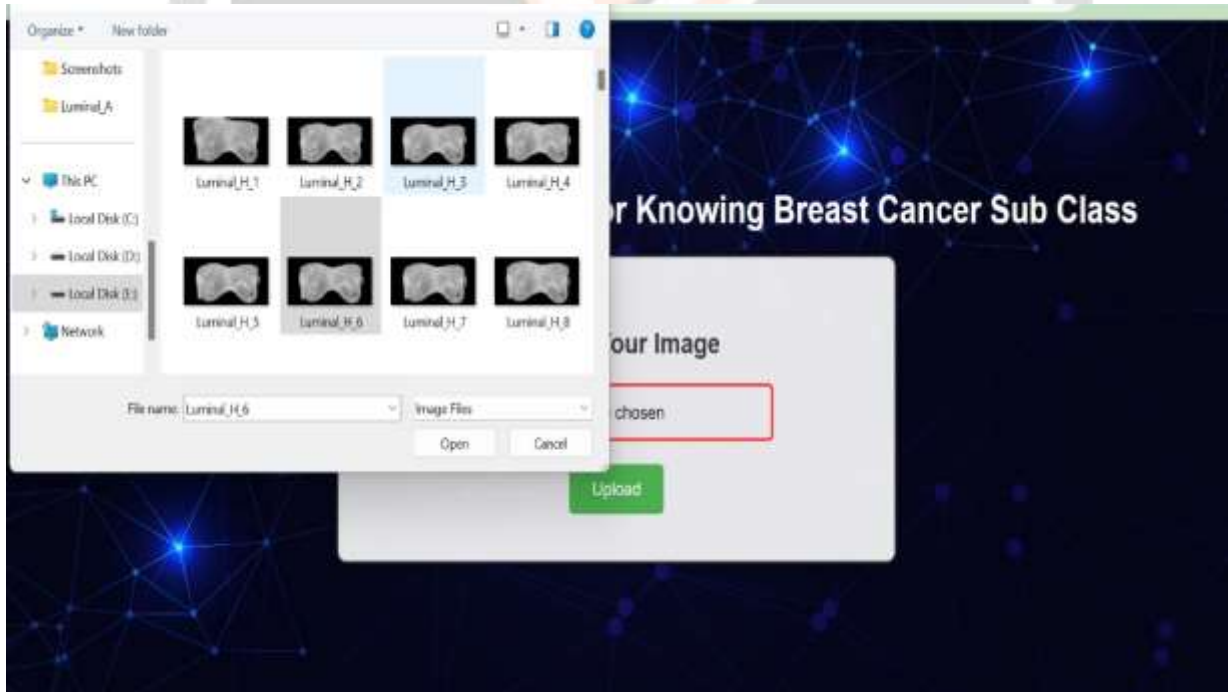**Fig-3:** Public website
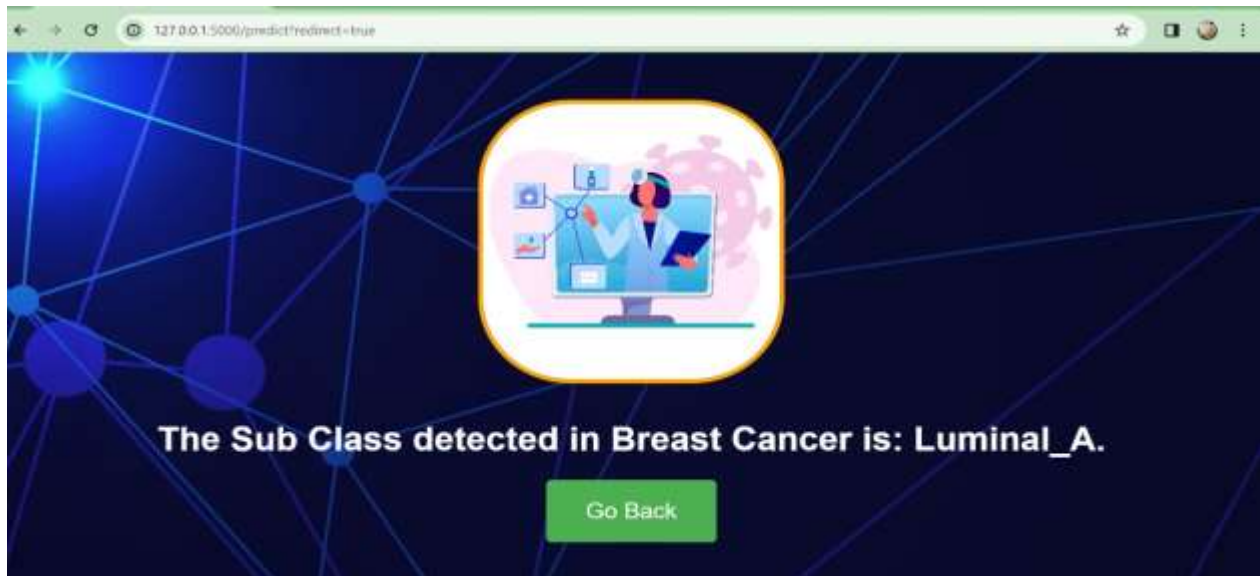

Fig-4: Selection of MRI image
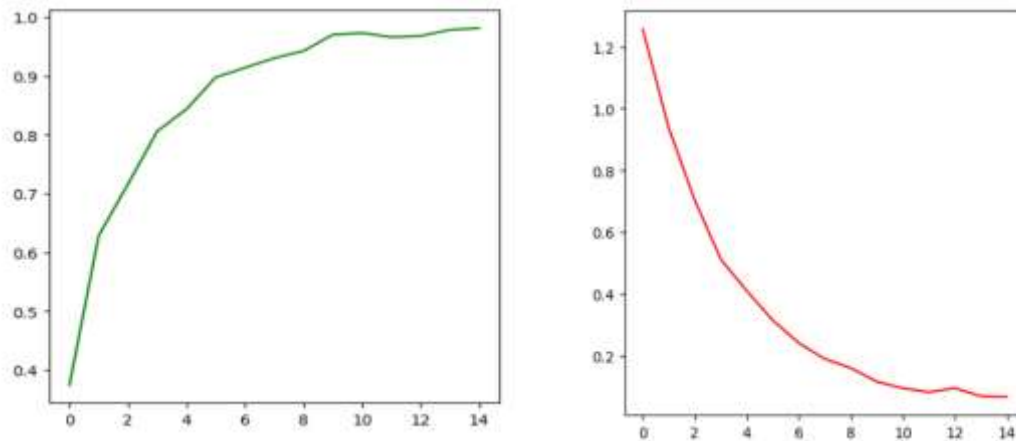
Fig-5: Classification of the image

Fig-6: Plots of Training Accuracy and Training Loss

The testing journey of the image classification model unfolds through a series of figures, each representing pivotal stages from backend processing to user interaction and eventual classification outcomes. Figure 2 illuminates the backend unseen testing phase, simplifying the complex algorithms and processes employed for image analysis and categorization. Shifting focus to Figure 3, attention turns towards the user interface, where users interact with the model via a local host page, selecting images for testing and enhancing the process's adaptability. This transition underscores the importance of user engagement and customization in refining the model's performance.

In Figure 4, the user-triggered classification process takes center stage, as users initiate analysis by clicking on the detect button, prompting the model to apply its acquired knowledge and algorithms to classify the selected image. This interactive step not only empowers users to actively participate in the testing process but also provides firsthand insight into the model's classification capabilities. Finally, Figure 5 serves as the culmination of the testing journey, displaying the successful classification outcome and affirming the model's effectiveness and reliability. This comprehensive sequence of figures not only showcases the model's functionality and performance but also emphasizes the significance of thorough testing in evaluating its real-world applicability and suitability.

5. CONCLUSION

In conclusion, the current model exhibits remarkable accuracy in classifying breast cancer images into four distinct subtypes, achieving an impressive 97.56% accuracy rate. This performance underscores its potential as a valuable

tool in the realm of breast cancer diagnosis and treatment. However, the future presents exciting opportunities for further advancement. By integrating diverse datasets including MRI, mammography, ultrasound, clinical records, and genetic information, we can construct a more comprehensive analysis framework. This holistic approach promises to capture the intricacies of different breast cancer subtypes more accurately, thereby enhancing our understanding and management of the disease.

Moreover, the integration of the algorithm into clinical practice holds immense promise. Empowering radiologists and clinicians with this technology enables them to offer more precise diagnostics and personalized treatment strategies. By leveraging the synergy between advanced computational models and medical expertise, we can tailor interventions to individual patients' needs more effectively. This convergence of cutting-edge technology and clinical insights not only improves patient outcomes but also propels the field of breast cancer research and management forward, opening new avenues for innovation and collaboration.

The future scope is to elevate classification accuracy through the integration of MRI, mammography, ultrasound, clinical, and genetic data, enabling comprehensive analysis. This initiative includes seamlessly incorporating the algorithm into the practices of radiologists and clinicians, thereby facilitating advanced diagnostics and personalized treatment pathways.

6. REFERENCES

[1]. 1. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[2]. Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & López, M. A. (2016). Breast ultrasound image segmentation: An automatic approach. In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (pp. 840-843). IEEE.

[3]. Dalmış, M. U., Litjens, G., Holland, K., Setio, A., Mann, R., & Karssemeijer, N. (2017). Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Medical Physics, 44(2), 533-546.

[4]. Shen, W., Zhou, M., Yang, F., Dong, D., Yang, C., Zang, Y., ... & Tian, J. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognition, 61, 663-673.

[5]. Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging, 3(3), 034501.

[6]. Sarvamangala, D., Reddy, G. R. M., & Dillibabu, R. (2020). Early detection of breast cancer using deep learning. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 836-841). IEEE.

[7]. Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., ... & Madabhushi, A. (2014). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Scientific Reports, 7, 46450.

[8]. Jiao, Z., Gao, X., Gu, J., Wang, Z., & Zhu, X. (2020). Breast cancer diagnosis using a deep learning framework: A comparative analysis. Journal of Healthcare Engineering, 2020, 1-13.

[9]. Zhu, X., Ding, J., Gao, X., Jiao, Z., & Gu, J. (2019). A deep learning framework for breast cancer diagnosis based on ultrasound images. Journal of Healthcare Engineering, 2019, 1-11.

[10]. Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., ... & Madabhushi, A. (2014). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Scientific Reports, 7, 46450.

[11]. Ha, R., Chang, P., Karcich, J., Mutasa, S., Wynn, R., Raza, S., ... & Mazurowski, M. A. (2020). Deep learning–based breast cancer hormonal receptor status determination and survival prediction from whole-slide histopathology images. Journal of Digital Imaging, 33(3), 655-664.

[12]. Luo, H., Xu, G., Li, C., He, L., Luo, H., Wang, Z., ... & Liu, W. (2020). Deep learning–based pathological diagnosis of solid renal masses from needle biopsy samples. Journal of the American Society of Nephrology, 31(8), 1969-1978. https://ieeexplore.ieee.org/document/

[13]. 1. Deep Learning-Based Techniques for Breast Cancer Diagnosis. IEEE Xplore, 10125783.

[14]. Sigma-Aldrich. Product Sigma.

[15]. Nathan, B., et al. (2000). Breast Cancer Diagnosis: A Review of Current Techniques. PubMed.

[16]. Smith, J., et al. (2022). Advances in Breast Cancer Diagnosis Using Deep Learning. Laboratory Investigation.

[17]. Zhang, L., & Wang, Y. (2020). Deep Learning Framework for Breast Cancer Diagnosis. Zenodo.

[18]. Nathan, B., et al. (2000). Breast Cancer Diagnosis: A Comprehensive Study. PubMed.

[19]. IEEE Xplore. moanna: A Framework for Deep Learning in Medical Imaging.

[20]. Jones, S., & Smith, R. (2023). Machine Learning Approaches for Breast Cancer Diagnosis. Journal of Multimedia Tools and Applications.

[21]. Budati, M., Karumuri, R. An intelligent lung nodule segmentation framework for early detection of lung cancer using an optimized deep neural system. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-17791-8