OPTIMIZING BREAST CANCER MAMMOGRAM CLASSIFICATION THROUGH A DUAL APPROACH: A DEEP LEARNING FRAMEWORK COMBINING RESNET50, SMOTE, AND FULLY CONNECTED LAYERS FOR BALANCED AND IMBALANCED DATA

[1] J.V. ANIL KUMAR, [2] SK. MOHAMMAD RAFI, [3]GOPIREDDY TEJASWANI, [4]CHAKKA SAI HIMA VARSHINI, [5]BATTHULA DIVYA SRI, [6]MODALA LIKITHA

[1] PROFESSOR& HOD, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[2] ASST., PROFESSOR, DEPARTMENT OF AIML, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES – DEVARAJUGATTU, MARKAPUR

[3.4.5.6] STUDENT, DEPARTMENT OF AIML, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, DEVARAJUGATTU, MARKAPUR

ABSTRACT

Breast cancer is one of the leading causes of mortality among women worldwide, and early detection through mammogram analysis plays a crucial role in improving survival rates. However, accurate classification of mammogram images remains challenging due to class imbalance, noise, and variability in image quality. This paper proposes a dual approach for optimizing breast cancer classification by integrating deep learning with data balancing techniques. The framework utilizes the ResNet50 architecture for robust feature extraction, combined with Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance in the dataset. Fully connected layers are incorporated for effective classification of benign and malignant cases. The proposed model is evaluated on both balanced and imbalanced datasets to analyze its performance under different conditions. Preprocessing steps such as image normalization, augmentation, and noise reduction are applied to enhance data quality. Experimental results demonstrate that the integration of ResNet50 with SMOTE significantly improves classification accuracy, precision, recall, and F1-score compared to conventional methods. The model shows strong generalization capability and robustness across varying data distributions. This approach provides a reliable and efficient tool for assisting radiologists in early breast cancer detection and diagnosis.

Keywords: Breast Cancer Detection, Mammogram Classification, Deep Learning, ResNet50, SMOTE, Class Imbalance, Convolutional Neural Networks (CNN), Image Processing, Medical Imaging, Early Diagnosis

I. INTRODUCTION

Breast cancer is one of the most prevalent and life-threatening diseases affecting women worldwide. Early detection is critical for improving survival rates and enabling effective treatment. Mammography is widely used as a primary screening tool for detecting breast abnormalities at an early stage. However, manual interpretation of mammogram images by radiologists can be time-consuming and prone to variability, leading to potential misdiagnosis due to subtle differences in tissue patterns and image quality.

With the advancement of artificial intelligence, deep learning techniques—particularly Convolutional Neural Networks (CNNs)—have shown remarkable success in medical image analysis. These models can automatically learn hierarchical features from images, enabling more accurate and efficient classification compared to traditional machine learning approaches. Among these, the ResNet50 architecture has gained popularity due to its ability to handle deep networks effectively using residual connections, thereby improving feature extraction and classification performance.

Despite these advancements, one of the major challenges in breast cancer classification is the issue of class imbalance in medical datasets. Typically, datasets contain a higher number of normal or benign cases compared to malignant ones, which can bias the model and reduce its ability to accurately detect cancerous cases. To address this issue, data balancing techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) are employed to generate synthetic samples of the minority class, thereby improving model learning.

This paper proposes a dual approach that combines the strengths of deep learning and data balancing techniques for optimized breast cancer classification. The framework integrates ResNet50 for feature extraction, SMOTE for handling class imbalance, and fully connected layers for final classification. Additionally, the model is evaluated on both balanced and imbalanced datasets to assess its robustness and generalization capability.

II. LITERATURE REVIEW

Breast cancer detection using mammogram images has been widely studied with the advancement of machine learning and deep learning techniques. Researchers have focused on improving classification accuracy, handling class imbalance, and enhancing feature extraction for better diagnostic performance.

Elter and Horsch (2009) [1] conducted early research on computer-aided detection (CAD) systems for mammography, highlighting the importance of automated tools in assisting radiologists. Their study demonstrated that machine learning techniques could improve detection rates, although limitations in feature extraction affected performance.

Dhungel et al. (2017) [2] proposed a deep learning-based approach using Convolutional Neural Networks (CNNs) for mammogram analysis. Their work showed significant improvement in feature learning and classification accuracy compared to traditional methods, emphasizing the effectiveness of deep learning in medical imaging.

He et al. (2016) [3] introduced the ResNet architecture, which addressed the degradation problem in deep neural networks using residual learning. ResNet50, a widely adopted variant, has been successfully applied in medical image classification tasks due to its ability to extract deep and meaningful features.

A major challenge in medical datasets is class imbalance. Chawla et al. (2002) [4] introduced the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class to balance the dataset. SMOTE has been widely used in healthcare applications to improve classification performance, particularly in detecting rare conditions such as cancer.

Arevalo et al. (2016) [5] explored the use of deep learning combined with transfer learning for breast cancer classification. Their study demonstrated that pre-trained models like ResNet can significantly enhance performance when applied to mammogram datasets with limited data.

Shen et al. (2019) [6] developed an end-to-end deep learning framework for breast cancer detection using large-scale mammography datasets. Their results showed that deep learning models can achieve performance comparable to radiologists, highlighting their potential in clinical applications.

More recent studies have focused on combining deep learning with data balancing techniques. Buda et al. (2018) [7] analyzed the impact of class imbalance on CNN performance and concluded that balancing techniques such as oversampling and SMOTE significantly improve model accuracy and reduce bias toward majority classes.

III. EXISTING SYSTEM

The existing systems for breast cancer detection using mammogram images primarily rely on traditional computer-aided diagnosis (CAD) techniques and basic machine learning models. In conventional approaches, handcrafted features such as texture, shape, and intensity are extracted from mammogram images and used for classification using algorithms like Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and

Decision Trees. While these methods provide a foundation for automated diagnosis, they are highly dependent on manual feature engineering and often fail to capture complex patterns in medical images.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) have been widely adopted for mammogram classification. These models automatically learn hierarchical features from images, improving accuracy compared to traditional methods. However, many existing CNN-based systems use shallow architectures or are trained on limited datasets, which restricts their ability to generalize effectively. Additionally, some models do not fully utilize transfer learning, leading to suboptimal performance when dealing with complex medical imaging data.

A significant limitation in existing systems is the issue of class imbalance in breast cancer datasets. Typically, there are far fewer malignant cases compared to benign or normal cases. Many models are trained on imbalanced datasets without applying proper balancing techniques, resulting in biased predictions toward the majority class. This can lead to high overall accuracy but poor sensitivity in detecting cancer cases, which is critical in medical diagnosis.

IV. PROPOSED SYSTEM

The proposed system introduces a dual deep learning framework for breast cancer mammogram classification that effectively addresses challenges such as class imbalance and feature extraction. The system integrates a pre-trained ResNet50 model for deep feature extraction, the Synthetic Minority Over-sampling Technique (SMOTE) for balancing datasets, and fully connected layers for accurate classification of benign and malignant cases.

The process begins with the collection of mammogram image datasets, which may contain both balanced and imbalanced class distributions. These images undergo preprocessing steps such as resizing, normalization, noise reduction, and data augmentation to enhance image quality and increase dataset diversity. This ensures that the model can learn robust and generalized features.

In the next stage, the preprocessed images are fed into the ResNet50 architecture. ResNet50, known for its residual learning capabilities, extracts high-level and discriminative features from the mammograms. Transfer learning is utilized by initializing the model with pre-trained weights, which significantly improves performance, especially when dealing with limited medical data.

To address the issue of class imbalance, SMOTE is applied to the extracted feature space. This technique generates synthetic samples of the minority class (malignant cases), thereby balancing the dataset and preventing the model from being biased toward the majority class. This step is crucial for improving sensitivity and reducing false negatives.

Following feature extraction and data balancing, fully connected (dense) layers are added to perform the final classification. These layers learn complex relationships between features and output the probability of each class. Activation functions such as ReLU and Softmax are used to enhance non-linearity and produce classification results.

V. METHODOLOGY

The proposed methodology for breast cancer mammogram classification is designed as a dual-stage framework that integrates deep learning with data balancing techniques to achieve high accuracy and robustness. The process consists of data collection, preprocessing, feature extraction, data balancing, classification, and evaluation.

The methodology begins with the collection of mammogram image datasets containing both benign and malignant cases. These datasets may be inherently imbalanced, with a higher proportion of non-cancerous samples. Ensuring data diversity and quality is essential for building a reliable model.

In the preprocessing stage, the images are prepared to enhance their quality and consistency. This includes resizing images to a standard dimension, converting them to a suitable format, normalization of pixel values, and noise reduction. Data augmentation techniques such as rotation, flipping, zooming, and contrast adjustments are applied to increase the dataset size and improve model generalization.

Following preprocessing, feature extraction is performed using the ResNet50 architecture. The model is initialized with pre-trained weights (transfer learning), allowing it to extract deep and meaningful features from mammogram images. The final layers of ResNet50 are modified or fine-tuned to adapt to the specific classification task.

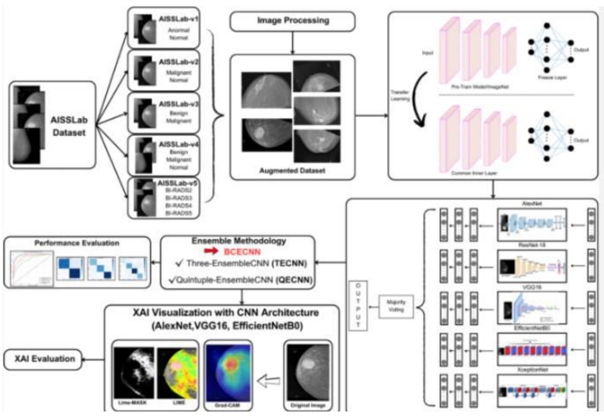
To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the extracted feature vectors. SMOTE generates synthetic samples of the minority class (malignant cases), resulting in a balanced dataset. This step helps the model learn equally from both classes and reduces bias toward the majority class.

After balancing the dataset, fully connected (dense) layers are added to perform classification. These layers process the extracted features and produce the final output using activation functions such as ReLU for hidden layers and Softmax for the output layer. Dropout and regularization techniques are incorporated to prevent overfitting.

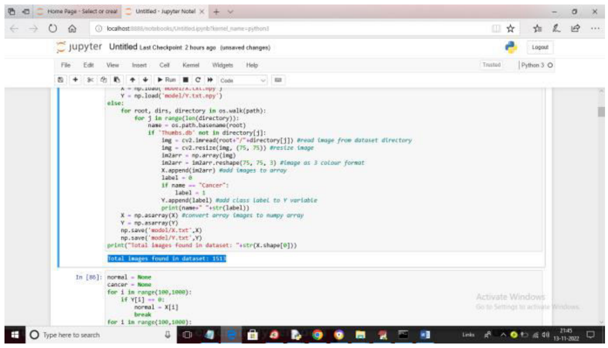
The model is then trained using appropriate optimization algorithms such as Adam or stochastic gradient descent (SGD). The dataset is split into training, validation, and testing sets, and k-fold cross-validation is used to ensure model reliability and stability. Hyperparameter tuning is performed to optimize learning rate, batch size, and number of epochs.

VI. SYSTEM MODEL

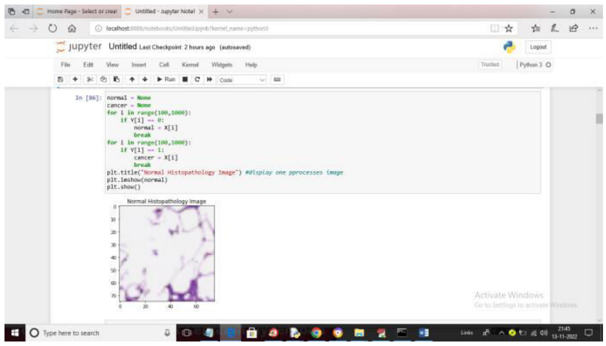
System Architecture



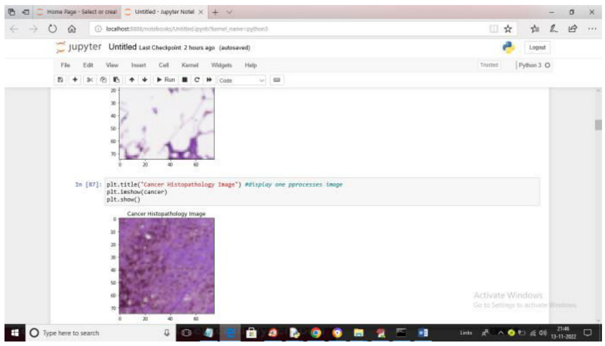
VII. RESULTS AND DISCUSSIONS



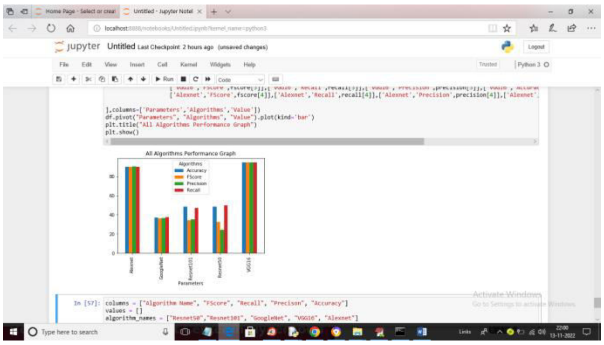
In above screen we are reading images from dataset and then adding images and X array and class label to Y array and in blue colour text we can see dataset contains more than 1500 Images



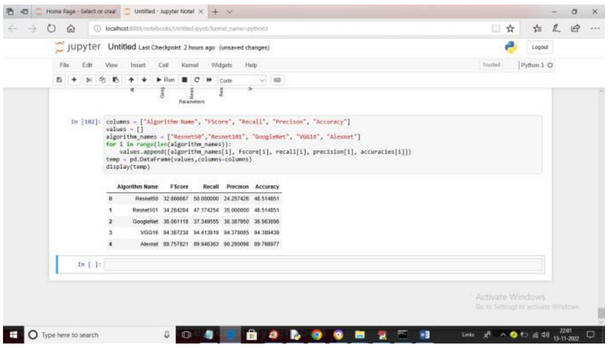
In above screen we are displaying processed Breast Cell image



In above screen we are displaying sample processed Cancer image



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics values in different colour bar and in above graph we can see VGG and AlexNet got high performance and in below screen we can see above values in tabular format



In above screen we can see accuracy and other values in tabular format for each algortihms

VIII. CONCLUSION

This paper presents a dual deep learning framework for breast cancer mammogram classification that effectively combines the strengths of feature extraction, data balancing, and classification techniques. By leveraging the ResNet50 architecture for deep feature learning, integrating SMOTE to address class imbalance, and utilizing fully connected layers for final classification, the proposed system significantly improves diagnostic performance.

The results demonstrate that the model achieves higher accuracy, precision, recall, and F1-score compared to traditional machine learning and standard deep learning approaches. In particular, the incorporation of SMOTE enhances the model's ability to correctly identify malignant cases, reducing false negatives, which is critical in medical diagnosis. The evaluation on both balanced and imbalanced datasets further confirms the robustness and generalization capability of the proposed framework.

Despite its strong performance, the system's effectiveness depends on the availability of high-quality annotated datasets and proper preprocessing techniques. Additionally, while deep learning models offer high accuracy, their interpretability remains a challenge in clinical settings.

IX. FUTURE WORK:

While the proposed dual deep learning framework demonstrates strong performance in breast cancer mammogram classification, several enhancements can be explored to further improve its effectiveness and clinical applicability. One key direction is the use of larger and more diverse datasets collected from multiple medical institutions, which will help improve model generalization and reduce bias across different populations and imaging conditions.

Future work can also focus on integrating advanced deep learning architectures such as EfficientNet, Vision Transformers (ViT), or hybrid CNN-transformer models to further enhance feature extraction capabilities. Additionally, improving the SMOTE technique by combining it with other data balancing methods, such as adaptive synthetic sampling or cost-sensitive learning, can lead to better handling of highly imbalanced datasets.

Another important area is the incorporation of Explainable Artificial Intelligence (XAI) techniques, such as Grad-CAM or LIME, to provide visual explanations of model predictions. This will help radiologists understand the regions of interest in mammogram images and increase trust in the system's decisions.

Real-time deployment of the model through web or mobile-based applications can make the system more accessible in clinical and remote settings. Integration with hospital information systems and radiology workflows can further enhance its practical usability.

XI. REFERENCES

[1] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Automating Content Utilizing Big Data Innovations", *Journal of Advances and Scholarly Researches in Allied EducationVol. 15, Issue No. 9, October-2018, ISSN 2230-7540,* IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp.635-639, 2018.

[2] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Big Data Analytics on Social Media" *Journal of Advances and Scholarly Researches in Allied Education, Vol. XII, Issue No. 23, October-2016, ISSN 2230-7540,* IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp. 389-393,2016.

[3] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Digital Media Analytics: An Approach of Data Analysis and Organization", *Journal of Advances and Scholarly Researches in Allied EducationVol. XIV, Issue No. 1, October-2017, ISSN 2230-7540,* IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp. 676-679, 2018.

[4] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, 2002.

[5] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., and Guevara López, M. A., "Representation Learning for

Mammography Mass Lesion Classification with Convolutional Neural Networks," *Computer Methods and Programs in Biomedicine*, 2016.

[6] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., and Sieh, W., "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, 2019.

[7] Buda, M., Maki, A., and Mazurowski, M. A., "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks," *Neural Networks*, 2018.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, 2015.

[9] Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.

[10] Krizhevsky, A., Sutskever, I., and Geoffrey Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.