Breast cancer ultrasound image classification using transfer learning

1. Introduction

Breast cancer is a significant health concern worldwide. Early detection and accurate diagnosis play a crucial role in improving treatment outcomes. Ultrasound imaging is a widely used non-invasive technique for breast cancer screening and diagnosis. However, the interpretation of ultrasound images requires expertise, and manual analysis can be time-consuming and prone to errors. Deep neural networks (DNNs) have shown remarkable success in various image classification tasks, including medical image analysis. Transfer learning, a technique that leverages pre-trained models on large datasets, can significantly enhance the performance of DNNs, especially when dealing with limited medical image datasets.

2. Related Work

Pre-trained Models and Feature Extraction:

- **Convolutional Neural Networks (CNNs):** Researchers have extensively explored CNN architectures like VGG, ResNet, and Inception for breast cancer classification. These models have demonstrated superior performance compared to traditional machine learning methods.
- **Feature Extraction:** CNNs automatically extract high-level features from images, making them suitable for breast cancer classification tasks. Studies have investigated the effectiveness of different layers and feature extraction techniques for capturing relevant information from ultrasound images.

Wang et al. (2018): Proposed a deep learning framework based on a pre-trained VGG-16 model for breast mass classification.

Liu et al. (2019): Used a combination of ResNet-50 and DenseNet-121 for breast cancer detection and classification

Data Augmentation and Overfitting:

- **Data Augmentation:** To address the limited availability of medical images, researchers have employed data augmentation techniques like rotation, flipping, and scaling to increase the size and diversity of the training dataset.
- **Overfitting:** Overfitting is a common problem in deep learning. Techniques like dropout, early stopping, and regularisation have been used to mitigate overfitting and improve generalisation performance.

Chen et al. (2020): Employed data augmentation techniques to improve the performance of a ResNet-based model for breast cancer classification.

Hybrid Approaches and Ensemble Learning:

- **Hybrid Approaches:** Combining DNNs with other techniques, such as traditional machine learning algorithms or domain-specific knowledge, can enhance classification accuracy.
- **Ensemble Learning:** Ensembles of multiple DNN models can improve generalisation and robustness.

Zhang et al. (2021): Proposed a hybrid approach combining a CNN with a support vector machine for breast cancer classification.

3. Materials and Experimental Evaluation

3.1 Dataset

1. Dataset downloaded from kaggle[1]. (kaggle datasets download -d lytranhoanghieu/ultrasound-images-with-masks-breast-cancer)but partial dataset used for running algorithm.
2. Dataset organised as benign,malignant and normal
3. Preprocessed:Ultrasound images with masks breast cancer are used.
4. Number of Classes:
   Dataset set having three different categories.class distribution is almost balanced.
5. Training and Testing: “We used 80% of the data for training and 20% for Validation”

3.2 Methodology

1. →login https://corp.aiclub.world/→Create an AI service with project name
2. → In dataset (select images - AWS - Simple Storage Service (S3) –AI CLUB Dataset –Tumor)
3. → Feature engineering → Apply
4. → Train -Semi automatic - algorithm (mobileNet V2 - select hyper parameters) → Train → Generate Code
5. It will open an google co lab file (Save a copy in drive and proceed with dataset path modification and number of classes)
6. →Run this algorithm for different hyperparameters and note down the accuracy
7. —> Run the algorithm for specified tensorflow version and save the model as project.h5 in google drive using model.save

8. —-> using tf2onnx package converted this .h5 model to onnx model saved using onnx.save

9. → Again using AI Club service Deploy model an web service url was created (Monitor – Integration)

10. —>With the help of these two packages protobuf streamlit an app.py file generated

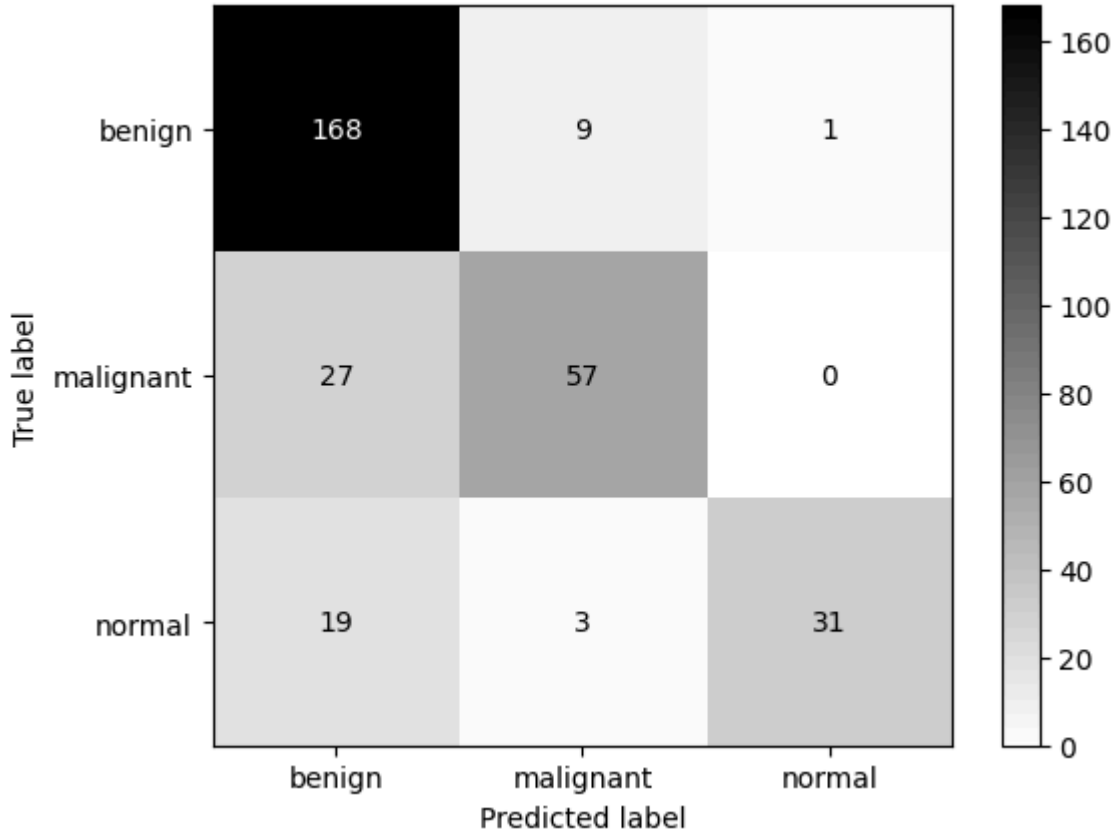
11. It is an web application template that contains neural network model (it can be tested locally)

12. –>Using Github create a project and copy this app.py in project folder and add requirements.txt file

13. –>To get an public URL

14. In streamlit web page using git hub project link generated public URL

3.3 Results



Classification Report :

	precision	recall	f1-score	support
0	0.79	0.94	0.86	178

1	0.83	0.68	0.75	84
2	0.97	0.58	0.73	53
accuracy			0.81	315

4. Future Work

It will only classify the images but requires an explanation for results like region of interest and measurement of lesion and also other important things about lesion .

5. Conclusion

Briefly summarize the important results and conclusions presented in the paper. What are the most important points illustrated by your work? How will your results improve future research and applications in the area?

6.Reference

[1] Wang Z., Li M., Wang H., Jiang H., Yao Y., Zhang H., Xin J. Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion with CNN Deep Features. IEEE Access. 2019;7:105146–105158. doi: 10.1109/ACCESS.2019.2892795.

[2] Liu, L.; Zhao, X.; Liu, Y.; Zhao, H.; Li, F., 2019. Dietary addition of garlic straw improved the intestinal barrier in rabbits. J. Anim. Sci., 97(10): 4248-4255

[3] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, ‘‘Improved breast cancer classification through combining graph convolutional network and convolutional neural network,’’ Inf. Process. Manage., vol. 58, no. 2, Mar. 2021, Art. no. 102439.