IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AUTOMATIC DETECTION AND CLASSIFICATION OF BREAST CANCER USING THE TRANSFER-LEARNING TECHNIQUE.

Dr.B.Geethavani[1],M.Tech,Ph.D, R.Gnaneswari[2], N. Hyndavi[3], V. Bharath Kumar[4], P.G. Hanumateja[5].

[1]Professor, [2345]UG Student [12345]Department of Computer Science and Engineering,
[12345]Siddharth Institute Of Engineering & Technology, Tirupathi, Andhra Pradesh, India.

ABSTRACT

Cancer is one of the most deadly diseases. In addition, breast cancer is the most common cancer in women and can be fatal if untreated. Despite substantial improvements in medical technology over the years, a biopsy is still the only effective way to find breast cancer. Pathologists can detect cancer using microscopic histology images. Visualizing cancer is a challenging task that requires a lot of skill, creativity, and focus. Hence, a quicker and more accurate way of detecting breast cancer is needed. As a result of advancements in machine learning and image processing, numerous sorts of research are being conducted to create an efficient partially or totally computer monitored diagnosing system. In this study, invasive ductal carcinoma was diagnosed and classified utilising

Keywords: *Deep learning, convolutional neural network (CNN), VGG19, Inception, Resnet, and dataset of breast cancer images.*

INTRODUCTION

The World Health Organization (WHO) reports that breast cancer is the most prevalent cancer in women worldwide. Three out of four impacted women will pass away. The current inquiry methodologies for breast cancer include mammography, magnetic resonance imaging (MRI), and pathology examinations. For patients who have already had additional investigations, such mammography, the histopathological photos are regarded as the gold standard to increase the accuracy of the diagnosis. Moreover, the histopathology analysis might offer more thorough and trustworthy Yongping Pan served as the assistant editor in charge of organising the evaluation of this article and giving final approval for publication. The laboratory technicians first use hematoxylin to stain the cell nuclei blue before counter-staining the cytoplasmic and non-nuclear components with eosin in various shades to highlight the various translucent tissue structures and cellular features in order to obtain the histopathological slides from the breast cancer tissues of the patients. The microscopic analysis of the stained biopsy tissues of breast cancer is then used to create digital histopathological pictures. Despite the fact that these images give pathologists (humans) a comprehensive look, mistakes can still occur if the diagnosis takes too long because of the large-sized slides. The number of women affected by breast cancer globally has sharply increased in recent years and is still rising. Lower death rates are achieved with early detection of

breast cancer. It puts off the onset of the illness until it becomes chronic. Around 1.7 million women will be affected by breast cancer in 2012. Every day, more people are impacted by breast cancer, making it more difficult for medical professionals to find them earlier and save their lives. Digital mammography is currently the primary method of diagnosis for locating breast cancer worldwide. Accurately diagnosing a wide range of illnesses typically involves computer-aided diagnosis (CAD). It aids medical professionals in recognising and classifying the various stages of sickness. CAD systems are designed to provide reassuring results and optimal assessments of the patient's condition, which helps medical personnel recognise the stages of diseases. It helps radiologists who use visually screening mammography of patients avoid erroneous assumptions and diagnoses brought on by incomplete information, a lack of attention, or inexperience. In order to save their priceless lives, the system aims to develop a ground-breaking CAD model that employs DNN to diagnose breast cancer in its early stages with more precise outcomes.

LITERATURE REVIEW

Deep Learning Architectures for H&E Histopathology Images Comparison

In many areas, including target recognition and image classification, deep learning has excelled. Recently, many research projects have concentrated on applying deep learning to the processing of medical images. The statistics of histological stains are known to be considerably different from the photographic RGB images in common deep learning imaging projects like as Imagenet and MIT Places, even though it is usual in image processing to utilise transfer learning for issues with small sample sizes. On the basis of data from histopathological staining with hematoxylin and eosin (H&E), this research assesses how well tuned models perform. We evaluate three convolutional neural network (CNN) designs in diverse contexts in order to analyse the performance of various deep learning architectures on these domains.

Finally, the effect of the size of the training sample context is assessed. We employ the BreaKHis dataset, which consists of microscopic images of breast cancer tissue stained in H&E. Our findings demonstrate that, in terms of accuracy and patient rate, fine-tuned architectures outperform neural networks that are trained from scratch.

Summary: In this paper, we evaluate the performance of transfer learning on H&E histopathology images. We conducted several experiments and came to the following conclusions: Transfer learning does have a satisfactory performance with a sufficient amount of contextual information, which can be adjusted by image resolution, crop size, and network selection to alter the first convolutional layer's kernel size. Finding a good fit between the kernel scale and the scale of the discriminating structures is the goal. Transfer learning provides an effective method for classifying and identifying H&E photographs. Better performance will be achieved by learning the entire network at a slower rate than the last fully linked layer. H&E images can also benefit from image statistics and frameworks that satisfy natural images. The performance and robustness of transfer learning will be enhanced by enhancing the training data and expanding the batch size and number of training samples. Compared to ImageData, training with lmdb input is substantially faster, and accuracy can be preserved.

HOW TO NORMALIZE HISTOLOGY SLIDES FOR QUANTITATIVE ANALYSIS

It is challenging to do a quantitative analysis on the results of histology slide preparation errors. The slides that were treated or kept under very varied conditions are brought into a single, normalised space in this work using two procedures that help to overcome many of the known discrepancies in the staining process. This enables more accurate quantitative analysis.

Summary: The research in this publication was conducted using slides of melanomas and nevi stained with haematoxylin and eosin, but it is transferable to other histological stains and tissues. On slides with various stain combinations, the algorithm for producing the best stain vectors has been satisfactorily tested. Results can occasionally be unpredictable when a slide contains three or more stains. The techniques in this study have significantly improved and have enhanced our investigation's outcomes. Greater reproducibility and support for larger datasets are two benefits of automating the procedure over manual selection techniques.

Breast cancer Histology Picture Classification Making Use Of Transfer Learning

One of the top causes of death for women is breast cancer. The improvement of survival rates, which have been rising gradually in recent years as a result of more powerful computer-aided-diagnosis (CAD) systems, depends on early detection and treatment. To minimise subjectivity and support the studies carried out by experts, CAD systems are crucial. For the goal of classifying breast histology image data into four tissue subtypes—normal, benign, in situ cancer, and invasive carcinoma—we suggest a transfer learning-based strategy. the photos from histology, In order to account for colour discrepancies brought on by slide preparation, the data supplied as part of the BACH 2018 grand challenge were first normalised. The Inception-V3 and ResNet50 convolutional neural networks (CNNs) from Google, which had already been trained on the ImageNet database, were then fine-tuned using the picture patches to help them learn the domain-specific characteristics required to categorise the histology images.

Summary:

This work presents a transfer learning-based strategy for categorising breast cancer histology images. Google's Inception-V3 and residual network (ResNet50) architectures, which were previously trained on ImageNet, are used to teach the network new features. Four tissue classes—normal, benign, in situ carcinoma, and invasive carcinoma—have been identified in the data set of photos submitted for the BACH 2018 Grand Challenge. In all three 3-fold cross validation trials, we trained all the networks using 80% of the data set for training and validation, and we tested their effectiveness using the remaining 20% of the images. The suggested transfer-learning method for automatically classifying breast cancer histology images is straightforward, effective, and efficient. Despite having little training data, the analysed networks were able to apply ImageNet knowledge stored as convolutional features to the challenge of classifying histological images. A CNN network that was constantly trained from scratch fared worse than the residual network (ResNet50) and Google's Inception-V3.

Shallu, Rajesh Mehra, "Classification of Breast Cancer Histology Images: Training from Scratch or Transfer Learning"

By comparing three previously-trained networks to the fully-trained network on the histopathological imaging modality, we were able to demonstrate the effectiveness of transfer learning.VGG16, VGG19, and ResNet50 and analyzed their behavior for magnification independent breast cancer classification. Concurrently, we examined the effect of training-testing data size on the performance of considered networks. A fine-tuned pre-trained With 92.60% accuracy, 95.65% area under the ROC curve (AUC), and 95.95% accuracy precision score (APS) for 90%–10% training–testing data splitting, VGG16 with logistic regression classifier produced the best result. Future aspects of this study could include layer-wise fine-tuning and various weight initialization approaches.

Summary:

By using three pre-trained networks (VGG16, VGG19, and ResNet50) for full training and fine tuning, this study examines the possibility of knowledge transfer from natural to histopathology pictures. These pretrained networks have been employed as a feature generator for transfer learning, and the generated features have been used to train logistic regression classifiers.

• The fine-tuned pre-trained VGG16 with logistic regression classifier produced the greatest performance with 92.60% accuracy and 95.65% area under the ROC curve when compared to the fully trained network.

• Compared to a fully trained network, a fine-tuned pre-trained network is more resilient to the size of training data since its performance does not significantly suffer when the size of training data is decreased.

• Classification tasks that are biassed in favour of one particular class significantly reduce network performance. In order to perform well, a network should be impartial or equally sensitive to all classes.

RECENT WORKS

There aren't any algorithms in place right now that can quickly identify cancer tumours from a dataset of images using deep learning. The measurements gathered by specific machines or by medical specialists, which may be a more expensive and time-consuming process, are used to detect breast cancer.

Icreased losses and decreased precision are drawbacks.

By taking into account three pre-trained networks—VGG16, VGG19, and ResNet50—and analysing their behaviour for magnification independent breast cancer classification, Shallu, Rajesh Mehra demonstrated the ability of transfer learning in comparison with the fully-trained network on the histopathological imaging modality. In parallel, we looked at how the amount of the training-testing data affected the effectiveness of the networks under consideration. The greatest performance came from a pre-trained VGG16 classifier that had been fine-tuned using logistic regression, with 92.60% accuracy, 95.65% AUC, and 95.95% accuracy precision score (APS) for 90%–10% training–testing data splitting. Future aspects of this study could include layer-wise fine-tuning and various weight initialization approaches.

Xieet al. carried out two experiments. The first experiment used 7,909 histopathology photos of breast cancer from the BreakHis database with imbalanced classifications.

Data augmentation methods, such as turning and clockwise rotation, were used in the second trial. There were an equal number of classes, and the average accuracy increased from 97.90% to 99.79%. The increased accuracy of this model emphasises the value of data augmentation methods for expanding the database and balancing the classes. Moreover, the 40resolution was used to attain the study's greatest accuracy score (99.79%). In addition, Xieet al. employed other pre-processing methods as normalisation, reducing border, and saturation correction. The optimised pre-trained VGG16 with logistic regression classifier outperformed the fully trained network in terms of accuracy, area under the ROC curve (AUC), and accuracy precision score (APS) over a 90-10 training-testing cycle.

Different resolutions of the BreaKHis and TMA datasets were integrated by Jannesariet al. Following that, data augmentation techniques were used to balance the number of categories and subtypes. 16,846 photos were used as input in total for this investigation. Prior to using fine-tuning approaches, they used pre-processing techniques like color-distortion and normalising to examine all the model's layers. Many experiments using ResNet models, including ResNet-152, ResNet-101, and ResNet-50, were carried out in the same study. These tests produced accuracy scores of 98.70%, 98.40%, and 97.80%, respectively.

This study demonstrated the significance of the deep layers where ResNet-152 had the highest accuracy of all ResNet models by employing the same methodologies and databases.

According to Han et alstudy, .'s the CSDCNN model had an overall accuracy rating of 93.20%. involves pre-processing, data augmentation, and final layer fine-tuning. For eight unbalanced classes, Inception-V3 had the greatest accuracy score (90.28%) of all the Inception models.

A deep CNN was used by Ting et al. to classify BC-lesions. There were 1 input layer, 28 hidden layers, and 1 output layer in this network. The feature-wise-data-augmentation (FWDA) technique prevented over-fitting. For sensitivity, accuracy, and specificity, their suggested technique consecutively achieved 89.47%, 90.50%, and 90.71%, respectively.

Toçaret al. proposed the BreastNet, which was capable of extracting the most useful features from breast images. It was composed of convolutional, pooling, residual, and dense blocks. As its accuracy approached 98.80%, BreastNet outperformed AlexNet, VGG-16, and VGG-19 models. Abbas introduced a multi-layer DL classification architecture. Images of the breast's benign and cancerous areas. This network had four stages for obtaining invariant features, deep-invariant features transformation, and learning features for the final judgement. The MIAS data set was employed in [82], and its sensitivity, specificity, accuracy, and AUC all received scores of 92%, 84.2%, 91.5%, and 0.91, respectively.

Shaet al. provided a method for automatic detection and categorization of the malignant zone in breast images using the same data set. Their suggested approach was based on the grasshopper optimization algorithm and CNNs. The findings indicated that this suggested strategy may, in terms of sensitivity, specificity, and accuracy, achieve 96%, 93%, and 92%, respectively.

PROPOSED PROCEDURE

The model we are presenting here uses the classification method to categorise breast cancer. The use of Deep learning algorithms to automatically classify malignant tumours is being investigated in this work. After removing the features from the images in the dataset, we will utilise the Breast cancer picture dataset to train the CNN, inception, Resnet, and Vgg19 algorithms. After training, we may test by supplying the input image.

ADVANTAGES:

Low complexity, high performance, precise categorization, and simple identification

ARCHITECTURE

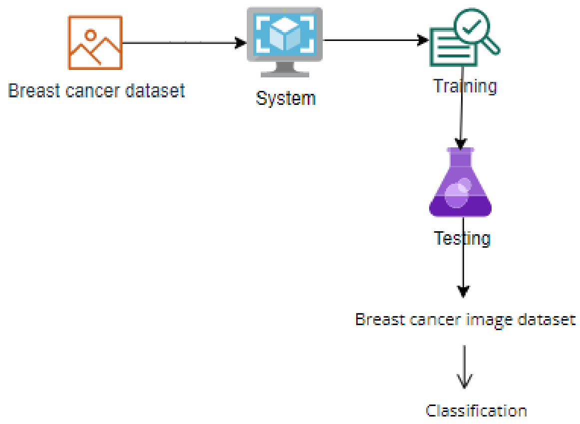

Fig 1: Architecture

ALGORITHMS AND METHOD

Neural Convolution Network

Step 1 : Convolution operation

The convolution operation is the first component of our strategy. We will discuss feature detectors in this phase because they essentially act as filters for neural networks. We'll also talk about feature maps, learning the specifications of such maps, the methods of pattern identification, the detection layers, and the mapping of the results.

Step (1b): Layer ReLU
The Rectified Linear Unit or Relook will be used in the second portion of this process. We will discuss Relook layers and examine the role of linearity in Convolutional Neural Networks. Although it's not necessary to comprehend CNN's, it wouldn't hurt to take a quick lesson to advance your knowledge.

Step 2: Conv2D
Keras Conv2D is a 2D convolution layer that helps construct a tensor of outputs by winding a convolution kernel with layers of input.
Kernel: A convolution matrix or mask is utilised in the image processing kernel and can be used for edge detection, sharpening, embossing, and blurring and more by convolution ally combining a picture with a kernel.

Step 3: flattening
Here is a basic explanation of the flattening procedure and how, when using convolutional neural networks, we get from pooling to flattened layers.

Stage 4: Complete Connectivity
Everything we discussed in the previous section will be combined in this section. You'll be able to imagine a more complete picture of how Convolutional Neural Networks function and how the "neurons" that are ultimately formed learn the classification of images by understanding how this works.
Finally, we'll put everything in perspective and provide a brief summary of the idea addressed in the section. If you think it will help you in any way (and it probably will), you should look at the additional as well as Cross-Entropy. Although it is not required for the course, it will benefit you greatly to be familiar with these principles since you will probably encounter them when working with convolutional neural networks.
An input layer, hidden layers, and an output layer are the components of a convolutional neural network (CNN). Any middle layers in a feed-forward neural network are referred to as hidden layers since the activation function and final convolution hide their inputs and outputs.

MODULES:
System:
1.1 Produce a dataset
The dataset including images of breast cancer that need to be categorised as normal is divided into training and testing datasets, with the test size being set at 30–20%.
1.2 Image resizing and reshaping as part of the pre-processing
1.3Training:
Use the pre-processed training dataset is used to train our model using CNN,Vgg19, Resnet and inception usingDeep learning algorithm along with some of the transfer learning methods.
1.4 Classification:
The results of our model are display of Breast cancerimages classification.

User:

2.1 Upload dataset
The user has to upload data which needs to be classified.
2.2View Results
The classified image resultsare viewed by user.

RESULTS ANALYSIS

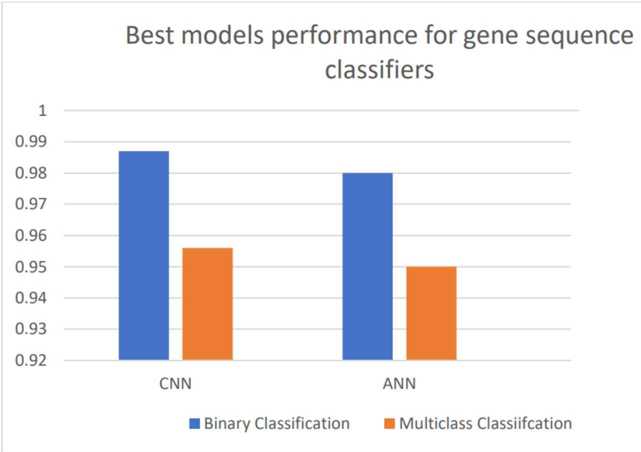

Fig 2: Results graph

Using deep learning and transfer learning, we effectively identified breast cancer photos in this experiment and categorized them according to whether they were impacted by the diseases associated with breast cancer. Here, we've taken into account the data set of breast cancer classification images, which will be of various sorts of breast cancer classification and trained using CNN.

CONCLUSION AND FUTURE SCOPE

In this experiment, we successfully recognised breast cancer photographs and categorised them according to whether they were affected by the diseases connected with breast cancer using deep learning and transfer learning. Here, we've taken into account the dataset of breast cancer classification images, which will be trained using CNN, VGG19, Inception, as well as some Resnet50 transfer learning approach, and will be of various types of breast cancer classification. After the training, we put the system to the test by uploading an image and classifying it.

In the future, this can be used to classify and detect the numerous illnesses and tumours that impact the human body, creating more accurate models that have a higher degree of dependability.

REFERENCES:

1. J. S. Hawley, The Bhakti Movement--from Where? Since When? India International Centre, 2009.
2. O. Linde and T. Lindeberg, “Object recognition using composed receptive field histograms of higher dimensionality,” in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004, vol. 2, pp. 1-6 Vol.2, doi: 10.1109/ICPR.2004.1333965.
3. Y. Ge and J. Yu, “A scene recognition algorithm based on covariance descriptor,” 2008 IEEE Int. Conf. Cybern. Intell. Syst. CIS 2008, pp. 838–842, 2008, doi: 10.1109/ICCIS.2008.4670816.
4. A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Contextbased vision system for place and object recognition,” Radiology, vol. 239, no. 1, p. 301, 2006, doi: 10.1148/radiol.2391051085.
5. H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: beyond empirical risk minimization, in: 6th International Conference on Learning Representations (ICLR) 2018, ICLR., 2018. URL: https://openreview.net/forum?id=r1Ddp1-Rb. Retrieved on 2021-04-28
6. M.H. Yap, N.D. Reeves, A. Boulton, S. Rajbhandari, D. Armstrong, A.G. Maiya, B. Najafi, E. Frank, J. Wu, Diabetic Foot Ulcers Grand Challenge 2020, 2020, https://doi.org/10.5281/zenodo.3715016.
7. X. Zhu, H. Hu, S. Lin, J. Dai, Deformable ConvNets V2: more deformable, better results, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 9300–9308, https://doi.org/10.1109/ CVPR.2019.00953.
8. Z. Cai, N. Vasconcelos, Cascade R-CNN: high quality object detection and instance segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 1483–1498, https://doi.org/10.1109/tpami.2019.2956516
9. J.H. Ng, M. Goyal, B. Hewitt, M.H. Yap, The effect of color constancy algorithms on semantic segmentation of skin lesions, in: Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics, SPIE, 2019, pp. 138–145, https://doi.org/ 10.1117/12.2512702

10. Y. Zhang, R. Liu, A. A. Heidari, X. Wang, Y. Chen, M. Wang, and H. Chen,``Towards augmented kernel extreme learning-models for bankruptcy prediction: Algorithmic behavior and comprehensive analysis,"Neurocomputing, vol. 430, pp. 185212, Mar. 2021.
11. J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang,C.-S. Huang, D. Shen, and C.-M. Chen, ``Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans," Sci. Rep., vol. 6, no. 1, pp. 113,Apr. 2016, doi: 10.1038/srep24454.
12. S. M. Ismail, A. B. Colclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. O'Sullivan, J. M. Summerell, and R. G. Newcombe, ``Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," BMJ, vol. 298, no. 6675, pp. 707710, Mar. 1989, doi: 10.1136/bmj.298.6675.707.
13. A. Andrion, C. Magnani, P. G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, and B. Terracini,``Malignant mesothelioma of the pleura: Interobserver variability," J. Clin. Pathol., vol. 48, no. 9, pp. 856860, Sep. 1995, doi: 10.1136/jcp.48.9.856.
14. J. Xie, R. Liu, J. Luttrell, and C. Zhang, ``Deep learning based analysis of histopathological images of breast cancer," Frontiers Genet., vol. 10, p. 80, Feb. 2019, doi: 10.3389/fgene.2019.00080.
15. H. Chen, Q. Dou, X. Wang, J. Qin, and P. Heng, ``Mitosis detection in breast cancer histology images via deep cascaded networks," in Proc. 13th AAAI Conf. Artif. Intell., Phoenix, AZ, USA, 2016, pp. 11601166.
16. M. Jannesariet al., ``Breast cancer histopathological image classification: A deep learning approach," in Proc IEEE Int. Conf.Bioinf. Biomed. (BIBM), Madrid, Spain, 2018, pp. 24052412, doi:10.1109/BIBM.2018.8621307.
17. MoloudAbdar,Mohammad Amin Fahami, Leonardo Rundo, Petia Radeva, Alejandro Frangi, U Rajendra Acharya, Abbas Khosravi "Hercules: Deep Hierarchical Attentive Multi-Level Fusion Model with Uncertainty Quantification for Medical Image Classification ",IEEE Transactions on Industrial Informics,vol.19,no.1,pp.274-285,2022.
18. Huan Liang, Wenlong Fu, Fengji Yi "A Survey of Recent Advances in Transfer Learning ",IEEE ICCT,pp.1516-1523,2019.
19. M.K. Nalini, K.R. Radhika " Comparative analysis of deep network models through transfer learning",Fourth International Conference on I-SMAC,pp.1007-1012,2020.
20. Seyyed Amir HadiMinoofam,AzamBastanfard" TRCLA: A Transfer Learning Approach to Reduce Negative Transfer for Cellular Learning Automata.",IEEE Transactions on Neural Networks and learning Systems,vol.10,pp.1109-1125,2021.
21. Haobin Shi, Jingchen Li, Jiahui Mao, Kao-Shing Hwang " Lateral Transfer Learning for Multiagent Reinforcement Learning. ",IEEE Transactions on cybernetics,2022.
22. Honggui Han, Hongxu Liu, Cuili Yang, JunfeiQiao "Transfer Learning Algorithm With Knowledge Division Level ",IEEE Transactions on Neural Networks and learning Systems ,vol.22,2021.