Histopathological analyses of breast cancer using deep learning

C. Ravindra Murthy[1*], K. Balaji[2]

[1]Research Scholar, Department of Electrical Electronics Engineering, St.Peter's Institute of Higher Education and Research, Avadi, Chennai
[2]Associate Professor, Department of Electrical Electronics Engineering, St.Peter's Institute of Higher Education and Research, Avadi, Chennai

*Corresponding author:
ravins.ch@gmail.com

Abstract

Deep Learning hosts a plethora of variants and models in Convolution Neural Networks (CNN), where the prudence of these methods is algorithmically proven when implemented with sturdy datasets. Much number of haphazard structures and textures are found in the histopathological images of breast cancer, where dealing with such multicolor and multi-structure components in the images is a challenging task. Working with such data in wet labs proves clinically consistent results, but added with the computational models will improvise them empirically. In this paper, we proposed a model to diagnose breast cancer using raw images of breast cancer with different resolutions, irrespective of the structures and textures. The floating image is mapped with the healthy reference image and examined using different statistics such as cross correlations and phase correlations. Experiments are carried out with the aim of establishing the optimal performance on histopathological images. The model attained satisfactory results and are proved good for decision making in cancer diagnosis.

Keywords

Histopathological images, Breast cancer, Deep learning, Floating image, Convolution Neural Network, Automated diagnosis

1 Introduction

Most developed countries has reduced death rate due to breast cancer in recent years, though there is annoying situations rise for massive growth of breast cancer. Technological break-through and awareness are the notifications observed by multitude prevent falling into the clutches of the epidemic, subsuming knowledge about medical imaging and analysis methods of early detection and screening. A meticulous scrutiny and diagnostic tests on patients shall have to be carried out with high sensitivity and specificity, in order to classify as 'cancer' or 'no-cancer'.

Research and Literary consensus demonstrate multi-level and multiple levels of resolutions on the histopathological images. Many time-hard analyses in the research jot on to valuable academic theses. In vitro data sets are the golden collection of the benchmarking data sets, where experts with minimal diagnostic risks in Computer Aided Diagnosis (CAD) has derived many sophisticated methods for decision making. Experts and research multitude annotate prospective decision making in building automated systems with promising features of reducing the risks and improving efficiency in diagnosis.

2 Literature Review

Classification is the process of categorizing and identifying the types of breast cancer from high-resolution histopathological images, used in deep learning using convolutional neural networks. Classifying benign and malign tissues is the order of work in the prognostic evaluation [1]. The carcinoma, which is also malignity in the tissues are studied with many morphological properties to determine the various levels of severities [2]. H&E (Heamatoxylin and Eosin) is the mechanism conventionally used to trace the primary stains on the specimens. Rule based methods in machine learning to CNNs of deep learning work with many digital compositions on the histopathological analyses, in order to achieve the automated end-to-end process [3].

Gradient descent and gradient boosting algorithms employed in the priority during the iterations of a CNN model in the Image Net [4]. Collections of H&E stained images of histopathological breast cancer forms are processed in the pre-trained ImageNet. A fully automated and operated CAD system uses deep learning and machine learning for the effective

diagnosis of breast cancer [5]. Detecting and extracting patches containing masses and classification is the core of the CNN frameworks. Early screening and detection, preprocessing of pectoral muscle by removal or mass segments examination need not encouraged, with external interventions, if CAD can do effectively.

Many inherent features and relationships in the histopathological images are discovered using machine learning. The performance of machine learning is further extended to convolutional neural networks, as prudent as hallmark for intelligent decision making. Learning based critical problems are solved meticulously using convolutional neural networks particular on the datasets with temporal and spatial aspects [6].

In traditional histopathological systems, there exists a consideration for the inter-pathologist-variability in all diagnostic experiments, which are well mitigated with quantitative analyses that consume much time in a clinical setup [7]. H&E promotes the aspects of variability for the stained areas, discussed as one of the major contributions of the research interests towards proving accuracy in image analysis.

Resolution, format and the structure of images influence the development of imaging devices, while large images occupy much storage space format and structure of the image plays a very important role. Whole-Side digital pathology Images (WSI) are the examples of the detailed imagery of histopathological breast cancer images [8]. These images are difficult to be sliced up into hundreds and many smaller tiles during the stage of preprocessing and selection, hence lowest and detailed magnification of imagery are considered for the tasks of classification and segmentation [9]. Metadata plays are very important role in sliced images to map with the positions of WSI, which need to be integrated with information. The experiments on the WSI is to break the barriers of handling errors during the mapping of sliced images with the whole format of the image, subsequently to provide as input to the convolutional neural networks to alleviate the challenges of uncertainty, size and format [10]. As problems of tumor identification are well dealt with machine learning frameworks, AI is the carrier for mitigating with the imaging problems and their sizes without having to compromise in the quality of the problem identification and description about the tumor detection, quoted by Robertson et al.

Histopathological images are used in medical domain such as oral cancer, Bone marrow, immune histo-chemistry [11]. They use some form machine learning approaches for the diagnosis from conventional methods to recent deep learning. An improved approach based on clump splitting is discussed in for breast cancer diagnosis [12]. At first, the nucleuses are segmented using clustering approaches and then classified using distance measure. Different shape features are extracted from the segmented nucleuses for the classification. A methodological review of histopathological image based breast cancer is discussed in [13]. The recent trends and methodologies are discussed along with the various challenged for the classification. An explanation method based breast cancer diagnosis is discussed in [14]. Three types of biases in the binary classification are discussed including sampling biases, biases correlated with class labels and biases those affect the entire dataset. A CNN based approach is discussed in [15]. It is a binary classification (benign/malignant) system which uses CNN for feature extraction and artificial neural network for classification. An ensemble approach is described in using different deep learning models. Four different models of visual geometry group architectures are utilized [16].

3 Proposed Method

Development of a software system for CAD, a plethora of datasets exists. These benchmark datasets are both useful in deep learning and conventional models. Benchmark datasets like Break His, have clinically relevant public breast cancer histopathology data set, with different kinds of trade-offs for practitioners, which is a very important study to date, however the relative availability of the clinical data should be compared with the benchmark data. The automated examination of pathological specimens is required for the diagnosis which runs in less time and cuts cost of analysis. Figure 1 shows the breast cancer classification according to the recommendations of WHO.

Histopathological Categorization of Breast Cancer	
Benign	Malign
Adenosis (A)	(DC) Ductal Carcinoma
Fibroadenoma (F)	(LC) Lobular Carcinoma
Phyllodes Tumor (PT)	(MC) Mucinous Carcinoma
Tubular Adenoma (TA)	(PC) Papillary Carcinoma

Figure 1. Breast Cancer Classified according to the recommendations of WHO.

Various resolutions of images and WSI have been considered for the experimentations, based on the degree of magnification of the image, the images are collected into four major categories of benign and malign class. The preprocessing stage of breast cancer diagnosis using convolutional neural networks begins with differentiating the images as benign or malign, further they are sub-categorized as shown above. For effective classification and sub-classification of images, to overcome the problem of iterative re-works, a binary classifier is employed to classify the identity and benign or malign.

The CNN which shall be employed on the data sets is based on the classification requirements, a binary classification can ideally work on magnification-specific training scenarios and multi-category classification works linearly on the training strategies, as magnification-independent. It requires training as four different models on different sets of specifically-magnified image data sets, which are as per WHO identification of benign and malign cancers. Data abstraction is considered during the implementation of the magnification-independent as all the levels of magnification are considered, unlike in the magnification-specific method.

4 Feature Extraction

Robust features of images are extracted by the methods evolved in the erstwhile research. Generally constituent blocks of images that reflect features are edge, corners, blobs, clouds and ridges. A particular property of the feature can be considered for the analyses in a particular domain, where the specific attributes as features. Properties are reflected as pixels representing the attributes into features of the physical view of the image. The scientific application of deep learning such as color-histograms, matrix based methods, binary pattern methods could be used systematically to extract the features. Computations and algorithms are used to extract the properties for the pathological studies, which may compete with the procedures of the experimentation in wet lab.

An image is floated considering the tumor/cancer, which is mapped with a healthy reference image, where the region of interest areas of the images is examined using similarity computations and statistics. Methods of statistics such as cross correlations, phase correlations, image ratio uniformity and difference of squares; other methods of mutual information ex-

change for the selected areas of the populated features include in the complex computations. Mapping of feature points of the healthy image with the floating images are examined in the similarity process. Certain parameters are identified and studied algorithmically from the healthy to tumor and are considered as the conversion parameters or transformation parameters, with points of bifurcation, cross-over points. Figure 2 shows the basic image analysis framework.
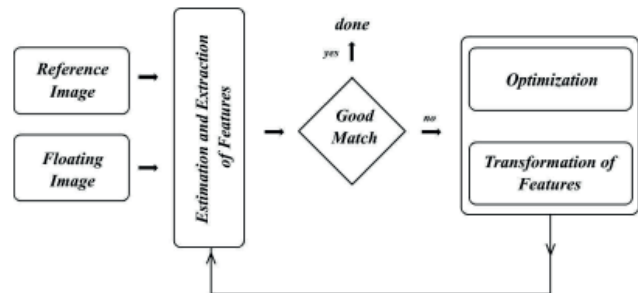

Figure 2. Basic Image Analysis Framework.

Cytoplasm or the nuclei pertaining to the breast cancer are not clearly visible components in the images, which may be due to grayscale aberrations. Tissue staining methods are implemented before visualization using sophisticated microscope which makes the procedure crucial to understand and highlight each component and helps in morphological analyses, where the similar objectives may also be achieved in wet lab.

4.1 Deep Learning Methodology

For the classification, the candidate data set from the corpus of images is selected, by scaling the original images into different resolutions. The learning time is influenced by the process of scaling and phasing of the images with a probable least time and irrelevant portions of the images could be eliminated from the process of learning. Though grayscale images are useful but cause aberrations and conflicting brightness which intrude the entire process of identification and detection of the tumor parts especially the shape of the tumor is the only identification of its benign or malign nature. Further, the CNN framework is implemented on the collected benign and malign classes of images. Figure 3 shows the samples of benign and malignant breast lesion.

5 Experimental Results

Many important aspects of malignant nature proved pathologically from the breast cancer images are revealed using the convolutional neural network.
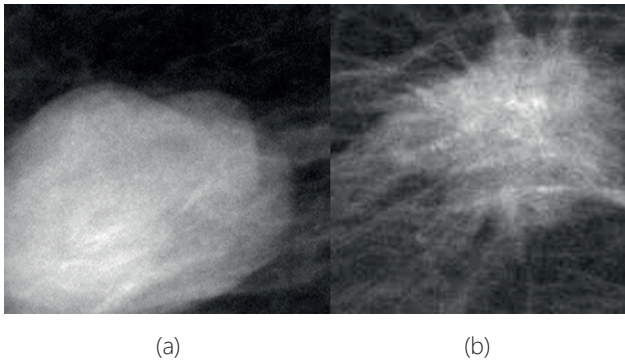
From the pashing and scaling the normalized images of breast cancer are sized to 220 x 220 are the primary tensors of size 192 x 192. Twenty four (24) filters with 3 x 3 x 2 kernel size are applied in the first convolution layer of configured CNN with a basic stride of 1 x 1. A max-pool with the stride of 2 x 2 is produced from the first convolution using the pooling layer reducing to 96 x 96. A ReLU is applied on the resulting output of the first convolution and sent through the subsequent convolution with nonlinearity and into the subsequent layers. For the second convolution operation kernel filters of size 3 x 3 x 24 with a total of 48 filtrations are applied and the input size reduced to 48 x 48, after max-pooling with stride of 2 x 2 the output is further scaled. To attain the nonlinearity, the output of the previous convolution layer is added to the current layer. For the third convolution operation kernel filters of size 3x3x96 with a total of 96 filtrations are applied and the input size is reduced to 24 x 24, after max-pooling with stride of 2 x 2 the output is further scaled. In the activation stage, ReLU function is employed, which further promotes to the fourth convolution layer. It has 192 filters with small kernel size of 3x3.

To clear the anomaly of activations, by filling the space during reduction of the output, the image is max-pooled with the stride of 12 x 12. Further convolutions are operated in order to process the results of all the pre-configured layers, which also include ReLU and 240 filtrations. The tensor 6 x 6 x 240 is result of the convolutions and then linearized and flattened to bring the shape of the feature. The values of the features within the neurons represent the patterns of the malign tissues.

CNNs face the problem of underfit and overfit. To overcome over fit, dropout layer is used, where the feature can be defined into the realistic format. Less number of neurons used to define the class of the datasets, in order to minimize the ambiguity from the fully connected layers. The fully connected layer at the final stage of convolution will bring out the tensor with limited number of neurons, in the experiment it is observed as 48, are converted to number of classes under the malign and benign. There will be a significant loss and while improvising accuracy occurred during the training and validation in the experimentation which are depicted in the following graphs. Figures 4 and 5 depict misclassified histological breast cancer photos, as well as the losses experienced during training and validation for accuracy in determining the benign and malignant character of images, respectively.

The generalization of the proposed model is based on the selection of the images, where the grading inaccuracies will affect the interest of deep learning methodology. From the observations of the experimentations conducted, the classification of breast cancer images as malign categories with their sub-categories, based on the selection of the datasets and propensity of the proposed method are referenced by the
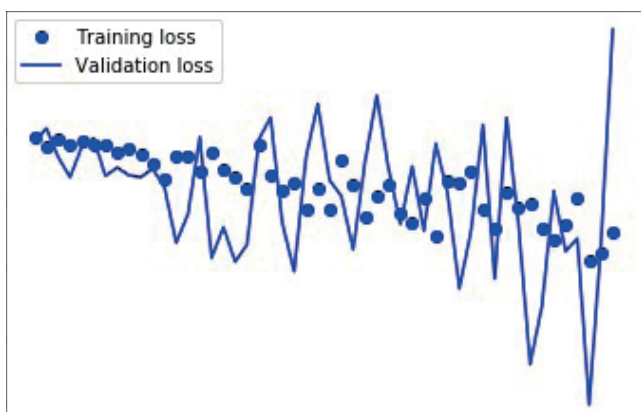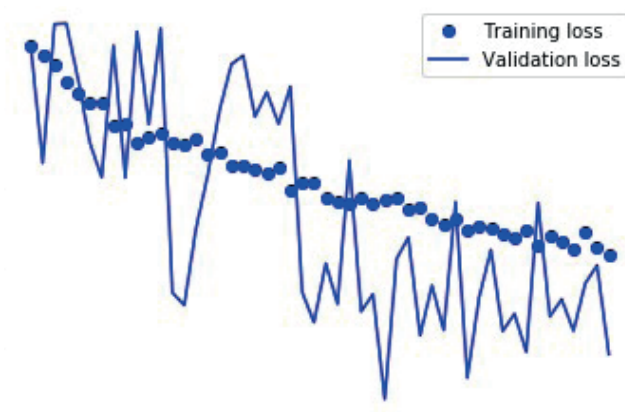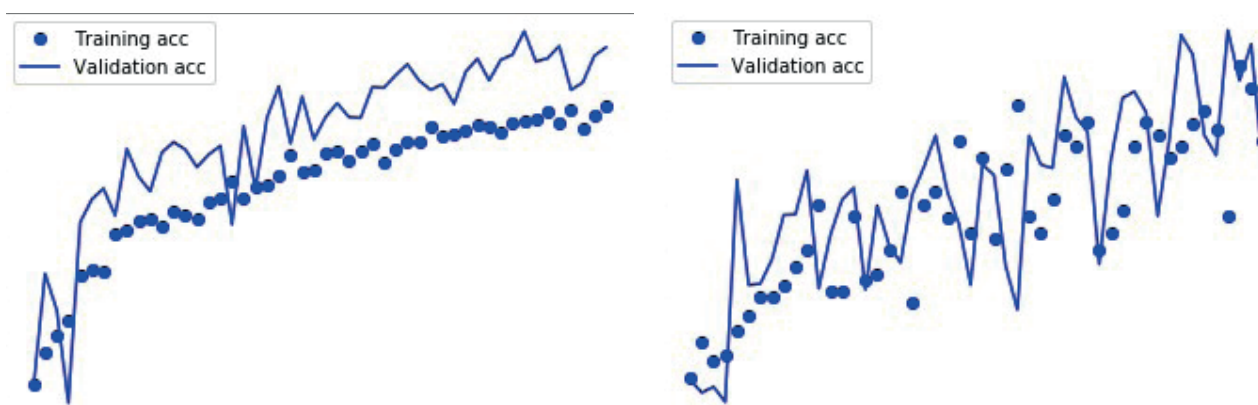
Figure 5. Misclassified histo-pathological breast cancer images and the loss incurred during training and validation towards accuracy of malign nature of images.

AUC of the ROC, which are drawn as follows determining for the malign and benign classes. Accuracy (ACC)=0.7866, Sensitivity (TPR)=0.7921, Specificity (TNR)=0.7837, False Positive Ratio (FPR)=0.2163, Positive Predictive value(PPV)=0.6597, Negative Predictive Value (NPV)=0.8769 are the ROC factors obtained for the benign class. Accuracy (ACC)=0.7849, Sensitivity (TPR)=0.788, Specificity (TNR)=0.7832, False Positive Ratio (FPR)=0.2168, Positive Predictive value (PPV)=0.673, Negative Predictive Value (NPV)=0.8671 are the ROC factors obtained for the malign class. Figure 6 shows the ROC curve which demonstrates the AUC for benign and malignant breast cancer images.

Therefore, the CNN proposes a learning model which has achieved measurable results on the scaled, phased and normalized histopathological images, with different dimensions and resolutions in the classification of the images as malign. Different kinds of CNN architectures may be proposed, which are almost like ImageNet, AlexNet etc,. Where the malign images of 240 x 240 are given as input to the convolution, pooling and ReLu of the CNN. The proposed model is prudent and flexible in deriving the desired results. Python with Tensorflow Keras in Anaconda Navigator with snippets of code in Jupyter Notebook and certain areas of adding activation function performed with colab of Google.

6 Conclusion

The CNN model for the classification of breast cancer using histopathological images has been proposed in the paper, which proves a simple CNN with a sequential model can be implemented for image

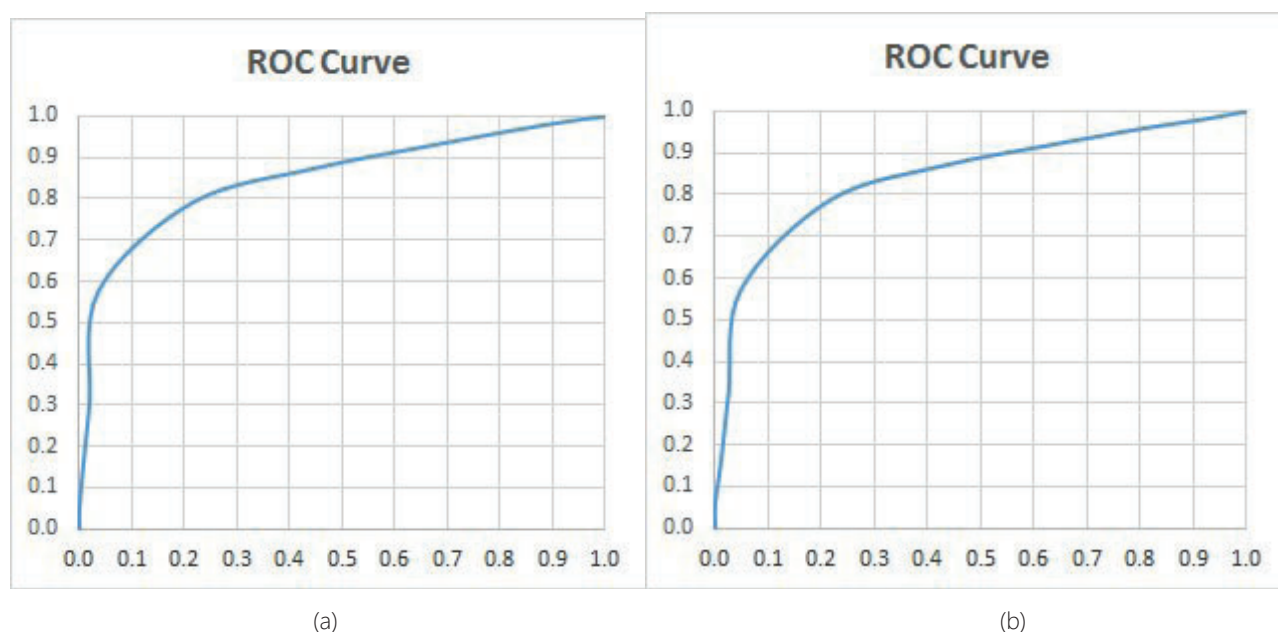(a)   (b)

Figure 6. ROC curve for (a) benign (b) malignant.

classification. It uses floating and reference images as inputs to the system and they are examined using similarity computations and statistics. To avoid the overfitting of CNN, dropout layer is implemented. All possible structures and textures of the features of the breast cancer can be thoroughly examined by overcoming various kinds of co color-scale aberrations. To illustrate the benefits of applying the propose method to breast cancer diagnosis, histopathological images are used. The proposed model has achieved the best results of classification of breast cancer images as malign. The proposed model can be applied in competition with the wet lab experiments, and has promising features of performing quantitative and qualitative analysis. More qualitative and good resolution images can be applied to obtain the absolute results of image classification.

Statement on ethical issues

Research involving people and/or animals is in full compliance with current national and international ethical standards.

Conflict of interest

None declared.

Author contributions

The authors read the ICMJE criteria for authorship and approved the final manuscript.

References

1. Elston, Christopher W., and Ian O. Ellis. "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up." Histopathology 19.5, Wiley Online Library (1991): 403-410.

2. Robertson, Stephanie, et al. "Digital image analysis in breast pathology—from image processing techniques to artificial intelligence." Translational Research 194, Elsevier (2018): 19-35.

3. Rakhlin, et al. "Deep convolutional neural networks for breast cancer histology image analysis", International conference image analysis and recognition, pp. 737- 744. Springer, Cham, 2018.

4. Spanhol, et al. "A dataset for breast cancer histopathological image classification." IEEE Transactions on Biomedical Engineering 63, No. 7 (2015): 1455-1462.

5. Shayma'a, et. al. "Breast cancer masses classification using deep convolutional neural networks and transfer learning." Multimedia Tools and Applications 79, no. 41 (2020): 30735-30768.

6. Khan et. al. "A survey of the recent architectures of deep convolutional neural networks." Artificial Intelligence Review 53, no. 8 (2020): 5455-5516.

7. Srinidhi et. al. "Deep neural network models for computational histopathology: A survey." Medical Image Analysis (2020): 101813.

8. Robertson, S et. al. "Digital image analysis in breast pathology—From image processing techniques to artificial intelligence". Transl. Res. 2018, 194, 19–35.

9. Bakare, Y.B. and Kumarasamy, M., 2021. Histopathological image analysis for oral cancer classification by support vector machine, International journal of advances in signal and image sciences, 7(2), pp.1-10.

10. Brück, O.E., Lallukka-Brück, S.E., Hohtari, H.R., Ianevski, A., Ebeling, F.T., Kovanen, P.E., Kytölä, S.I., Aittokallio, T.A., Ramos, P.M., Porkka, K.V. and Mustjoki, S.M., 2021. Machine learning of bone marrow histopathology identifies genetic and clinical determinants in patients with MDS. Blood cancer discovery, 2(3), p.238.

11. Lee, K., Lockhart, J.H., Xie, M., Chaudhary, R., Slebos, R.J., Flores, E.R., Chung, C.H. and Tan, A.C., 2021. Deep Learning of Histopathology Images at the Single Cell Level. Frontiers in Artificial Intelligence, p.137-148.

12. Ulle, A.R., Nagabhushan, T.N. and Manoli, N., Classification of Histopathological Images based on Modified Clump Splitting Approach. International Journal of Computer Applications, 975, pp.8887-8895.

13. Rashmi, R., Prasad, K. and Udupa, C.B.K., 2022. Breast histopathological image analysis using image processing techniques for diagnostic pupuses: A methodological review. Journal of Medical Systems, 46(1), pp.1-24.

14. Hägele, M., Seegerer, P., Lapuschkin, S., Bockmayr, M., Samek, W., Klauschen, F., Müller, K.R. and Binder, A., 2020. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. Scientific reports, 10(1), pp.1-12.

15. Dabeer, S., Khan, M.M. and Islam, S., 2019. Cancer diagnosis in histopathological image: CNN based approach. Informatics in Medicine Unlocked, 16, p.100231.

16. Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J. and María Vanegas, A., 2020. Breast cancer histopathology image classification using an ensemble of deep learning models. Sensors, 20(16), p.4373.