JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Detection and Classification of Oral Cancer Using Convolution Neural Network

Yasir Riyaz[1],Preeti Sondhi[2]

[1]M. Tech Scholar And Assistant Professor Department of Computer Science & Engineering, Universal Group of Institutions, Lalru, Punjab, India

Abstract:

Computerized classification of cell images into normal and malignant cells would be a vital aid in diagnostic processes, despite the fact that it has proven difficult to do. It has been shown that a number of photo classification systems benefit greatly from texture detection and processing. Dense Networks (DENSENETs), a texture-based approach that has demonstrated to have a lot of potential, have been used in recent papers. Some of these variants use DENSENETs-based convolutional neural networks (CNNs).. This project alters contemporary texture analysis. Three and two of the CNN structures used to identify images from a collection of both healthy and oral cancer cells are DENSENET-based. Results from Wieslander and Forslid using ResNet and VGG architectures as a benchmark, despite the fact that these designs weren't built with texture detection in mind. According to our study, DENSENET-Embedded CNNs perform better for certain task designs than traditional CNNs. Juefei-Xu ET's performance model outperformed the best reference model by 9 percent in F1-score and 0.5 percent in accuracy. It was 81.03 percent accurate and got an F1-score of 84.85 percent.

Keywords: CNN, LPB, VGG, Oral Cancer

I. INTRODUCTION

The early detection of cancer cells is essential for efforts to lower cancer mortality. Screening programs, such as those for cervical cancer, are effective in preventing malignancies in their advanced stages [12]. The cost of manually examining the resulting cell populations prevents the screening for additional cancer types, such oral cancer. Because of recent advancements in image processing techniques, these costs may be drastically lowered by allowing users to examine cytology slides. It is anticipated that methods focusing on texture analysis in particular will be useful for distinguishing between samples of healthy and cancerous cells..

Ojala demonstrated the effectiveness of dense networks (DENSENETs) [1] in conveying textures. Instead of focusing on the trend of intensity fluctuations with each pixel and a group of units by working closely with the picture amplitudes in the area around it, DENSENET. A neural network might then be trained to understand how these patterns are spread throughout various image classifications. This paradigm is investigated further.

A variety of techniques that either directly use DENSENETs or are impacted by them have been used to effectively classify a range of photo classification tasks, including the identification of common textures, the detection of Face spoofing, and the recognition of emotions. [6] gives a detailed explanation of these methods.

Convolutional neural network (CNN) models that use texture detection inside of spatial material characteristics are the most successful for the task of cell labeling..

This work updates and enhances three recently published image analysis techniques [3, 5, 8] that are suitable for categorizing and assessing textures for the detection of oral cancer. The results of these techniques are contrasted with those of Wieslander et al. [8], who employed two cutting-edge deep neural networks to analyze the same set: ResNet [1] and VGG [4]. Convolutional architectures are employed in two of the methodologies outlined in that are based on DENSENETs: LBCNN [3] and the "DENSEnet sum" model by [5]. The third strategy, named as RotEqNet [8], is used to construct a function with a schrödinger equation for texture categorization.

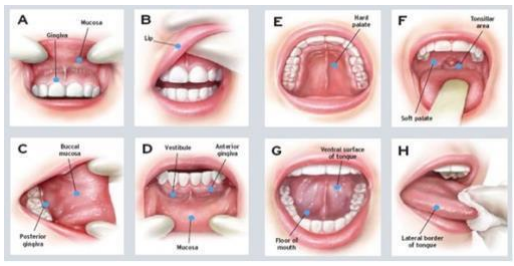


Figure 1: Probable oral cancer occurrence zones. .

II. LITERATURE REVIEW

AChodorowski et al [6] proposed a technique for classifying oral lesions using real color pictures. Five different color spaces were used to evaluate representations and their application to the processing of colonic images' color signals. Four well-known models were chosen for the categorization study: Sportfishing Sequential Substring, Gaussian polynomial, KNN, and the number of fully linked

layers are the first three. For the purpose of estimating classification accuracy, the reconstructing and fivefold cross validation procedures were employed. The best classification accuracy was attained using the HSI color space and the linear discriminant analysis classifier..

Woonggyu Jung et al. [8] proposed a method for the early detection of oral cancer using con-focal laser tomography. OCT is suitable for 2-4 mm level oral mucosa detection. Mouth cancer may be detected in 3-D scans of oral lesions. Research was done and a thorough study was written on the genetic perl diagnosis of oral cancer by Simon Kent [9]. He compared a neural network training approach with a neural network model..

III. METHODOLOGY

A. Local binary convolutional neural networks (LBCNN)

This approach uses modified Dense Network layers for classification syntheses. The authors note that a DENSENET index may be computed by applying P dense boolean filters in parallel, activate them using a Heaviside Function generator, and then producing a stacked total of the results with weights that are powers of two. The binary filters' only zeros would be the central pixel and one extra pixel. The total may be computed using a 1 1 P convoluted layer with weights v I = 2i for I = 0...P 1.

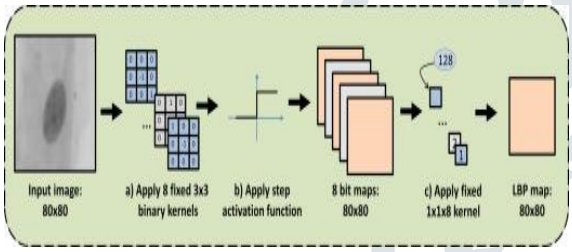


Figure 2: DENSENET design with convolutions

The DENSENET calculations in (a), carried out using the Heaviside linear system (b), are approximate for R = 1 and P = 8. The 1x1x8 kernel
(c) translates the mains adapter binary for each pixel on the input picture to create the related pattern identifiers between 0 and 255 with loads fixed at various powers of 2.

The new formula of DENSENETs as two fixed convolutions is seen in Figure 2.

The publications then broaden and change the framework in three distinct ways.:

1. 1. To begin, add up the weights of the layout elements. The pattern indices can only be used in a distribution since similar indices don't always show a clear trend. Instead of combining the P variances with different powers of 2 to produce a separate value for each possible motif, this problem
is avoided in step (c) by merging the difference maps in a weighted sum with memorization ratings for each point v0... vP1..

2. Resolved exponential or sigmoid action. Instead of binary outputs, radial basis stimulation is used in step for each map produced by the nonlinear activation step function (b). This metric's potential use in back propagation is due to its spatial separation..

3. Resetting filters at random intervals. P filters create the discrepancy maps among a core pixel and one neighbor that can be seen in step (a). Instead, screenings are randomly begun with 1 in a certain number of locations, leaving the remaining ones as zeros. A hidden layer, the end result of each filter, aggregates the contrasted and values of neighboring pixels..

4. . This results in a twofold soft-max layer of local pattern changes to the image pixel that is precisely the same size and shape. Therefore, these paired based classification convolutions (LBC units) provide a drop-in replacement for convolution operations in any model architecture.

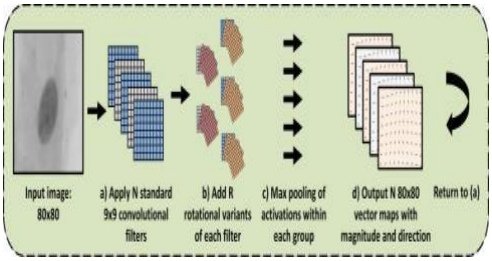


Figure 3: Rotational network that is reliable and effective (RotEqNet).

The authors granted: Each cie filter is the canonical filter for a group of circularly connected filters (b). The highest engagement value and the related spinning angle are the results of a voxel vector field (d). to train bigger, deeper nodes in the same time.

The decision to use just this one tactic was made because of the following reasons:

It is a semi-automatic process.

It is based on curve building and continues to find the boundary even when gradient specification is insufficient. It reduces demarcation time. It works correctly with all sample material despite the substantial amount of variability involved.

The planning is briefly outlined here.:

Step 1: To begin building the first mask, select the upper left and lower right positions.

Step 2: A typical surveying map (phi) is created starting with the mask.

Step 3: Select the narrow band of the curve.
Step 4: The various terms of the energy are computed. [3]

Stochastic gradient approach for energy reduction is the fifth step.

Step 6: The revised phi and curve propagation are computed.

Re-initializing phi to keep the Signed Distance Map smooth is step seven. [2]

Step 8: Before going back to step 3, display the intermediate output. Once the user-specified maximum number of iterations has been achieved, go on as before.

In step nine, a segmented ROI is overlay on the original image. 32x32 pixel pieces are selected, and they are saved as.tif files..

B. Patch Nomenclature

Each patch is assigned to one of the following assignees: group n gxxxass (n = normal, m = tumor)

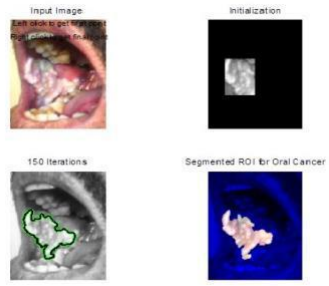


Figure 1. Database ROI Localization from Image Output GUI

Verrucous carcinoma in the buccal mucosa. the initial image (top left), the first mask (upper right), the boundary-filled halftones imagery after 150 repetitions (bottom left), and the split ROI for malignancy (lower right).

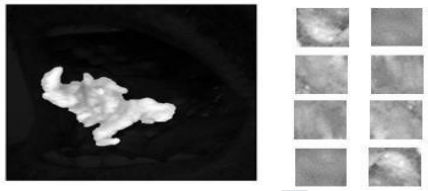


Figure 2: fractured ROI put on the original image to emphasize the greyscale ROI (left). Outstanding patches (right).

C. Patch Selection

Step 1: A segmented ROI is overlaid on the bottom image, as shown in Fig. 4, with the ROI displayed in grey and the rest of the image blacked.

Step 2: Using the ROI, choose flat pieces of 32x32 size by clicking the icon on as many different locations as required.
Step 3: Each patch, which is 32x32 pixels in size, is saved in.tif file using the language from Section 4.3.

Step 4: Repetition of steps 1 through 3 for every malignant image.

Steps 2 and 3 are finished for each typical photo..

D. Rotation Equivariant Vector Field

1. Networks (RotEqNet) [8]

The method described in [8] develops a brand-new vector field channel called RotEqNet in place of DENSENETs. This extends the earlier work by the same researchers ([9]), which proposes a single circularly invariant convolutional layer in which each filter is reproduced R times at different orientations. The scientists assert that the exceptionally large filters—measuring 35 by 35 pixels—will enable them to distinguish patterns and textures over a broader spectrum of wavelengths than tighter filters.

2. Utilizing a local binary pattern network, identify face faking [5]

Four building pieces make up this model: (1) a common convolution set, (2) a removing lbps module, (3) a producer for a programmable gate, and
(4) a set of common value of 1..

3. Convolutional Module

The first component of the broadcaster is a convolutional layer with N 3 3 filter dirt and phase leveling. Then comes a maxpool layer with stomp (2, 2). The final convolutional layer employs 2N 3 3 layers...

4. Gate Layer Module

The gate layer is made up of four successive sets of lenses that receive the out data from the DENSENET layer, similar to the structure of the DENSENET layer. In the input corpus filter, where y is the filter index, values between [y 1, y] are set to zero (1-8). Eighth nonlinear functions are produced as a consequence, with values ranging from [0, 1], [1, 2],... [7, 8] for filters 1, 2, and 8. A special operation is used once the equations have been filtered.

This method produced a normalizing proportionate sum of each cycle's numbers that increases to unity across the whole range [y1,y]..

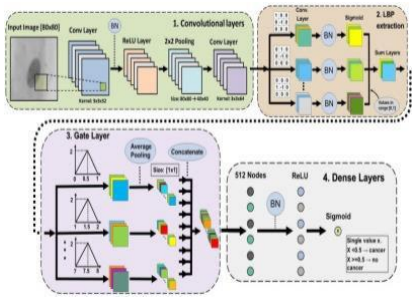


Figure 3: *Convolution layers*

The typical entire that follows reduces the size of each bit to 1x1. In our example, the gate layer's input layers had a 40 by 40 diameter, hence an aggregate pooling layer with a 40 by 40 kernel was used. The fourteen chained vectors are delivered to the density layer module. The outcomes of this approach may be conceptualized as the eight-bin processed spectrum of all the installations that have used DENSENET filters..

E. Data-set

The data set utilized in this study consisted of 10274 80 80 gray scale images of cells from six individuals, three of whom had cancer and three of whom did not. A tiny number of cells from the dental arches were taken, put on a glass, and photographed to produce a series of images utilizing the same cell sample. The images were then downsized such that one cell was visible in the center of each picture. Out of the 15 focus levels that were provided for each image, the variable that best fit the focus was determined by merging the image with the Laplacian
[8]. Figure 4 illustrates. Database ROI Localization from Image Output GUI.

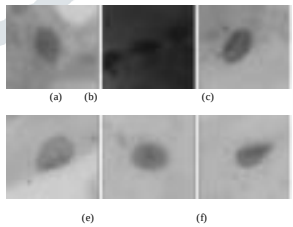


Figure 4. A patient in excellent health is shown in photos (a) through (c), and a subject with colorectal cancer is shown in pictures (d) through (f). Remember that the intensity of the images was adjusted prior to training

Table 1 displays how samples were divided between eyeglasses and doctors. The bottle number is provided because it should be taken into account when breaking up data for both training and assessment, given that it is likely that the images from one glass will be connected to one another. As a consequence, the data were split into three flips, as shown in Table 2. To ensure that both healthy and tumor cells are present in both the teaching and testing components and that no cells formed from learning eyeglasses are present in the assessment set, the entire dataset is chosen along with an evaluation set for each fold. Every time, a random 20% subset of the data set is chosen to act as the test datasets. .

Table 1 The oral cancer set includes people and lenses.

Patient	Glass	Diagnosis	Cells, size 80x80
1	3	Healthy	1965
1	4	Healthy	1851
2	5	Healthy	1382
2	6	Healthy	956
3	7	Healthy	863
3	8	Healthy	738
4	12	Cancer	226
5	36	Cancer	534
6	37	Cancer	963
6	38	Cancer	796

Table 2 Glasses are divided into several folds.

	Glasses for training		Glasses for evaluation	
	Healthy	Tumour	Healthy	Tumour
Fold 1	3,4,5,6	37,38,12	7,8	36
Fold 2	3,4,5,6	36,37,38	7,8	12
Fold 3	3,4,5,6	12,36	7,8	37,38

The classifiers are individually trained across each flip, and thus the final tally is created by adding the ambiguity indices per each run.

IV. EXPERIMENTAL STUDY

This part explains how and why the data were utilized, how the three network topologies mentioned in the introductory paragraph were modified to work best with our time series, and how they were improved for our experiments. The basic parameters from [3], [5], and [8] of the scientists were used as a starting point to alter loopy, including the number of filters in every layer, the length of layers, and the learning rate timeline..

A. Data Selection and Augmentation

By removing the mean and employing the standard deviations, each change is done to decrease the effects of any variances in ambient or dye circumstances across several materials. To standardize the training sets using as many pictures as possible, an equal amount of normal and sick photos are selected. Since each wave of this collection contains more positive data and cancer views, just a subset of the healthier photographs are included..

B. Local binary convolutional neural networks (LBCNN) [3]

We altered the Lua code provided by the authors of [2] in order to implement this strategy in Python. However, the algorithm was unable to train without it, so we added a second convolution layer but after the LBCNN level (in addition to batch normality before any LBCNN layer). Ten or twenty LBCNN layers with 128 auxiliary filters, each with 512 channels, were used, and the filter sparsities of 0.1, 0.5, and 0.9 were studied. Our averaging layer has a filter size of 5 by 5 instead of the 6 by 6 indicated in the source since our beginning picture size is larger than the publishers' design, producing output vector files that are appropriately 16 by 16 in size. The optimal results (shown in Table 3) were obtained with layers and a network configuration of 0.9 (90 percent of bits non-zero), with a typical totally know the material with 16 3 3 filter media inserted after the LBCNN levels and before fully linked layers.

C. Rotation Equivariant Vector Field 1. Networks

For this modeling, in addition to the creators' original Matlab code, we used PyTorch code supplied by Anders U. Waldeland at the Norwegian Computing Centre [15]. Three RotEqNet layers make up the implemented protocols. Because it was designed for images with a 32x32 pixel size and was accompanied by two entirely connected layers, we added a RotEqNet layer to make room for our 80 graphics. Furthermore, we found that the machine was using the learning process as planned, therefore we added a dropout's barrier with a rate of 0.7 after the RotEqNet layers..

The outcomes of this setup are shown in Table 3. One of the earlier numbers is the learning rate pattern, which starts at 0.1 and drops by a factor of ten at versions 20, 40, and 60. The tuning (batch trend line downhill withweight decay 0.01); the number of filter cycles used; and the channel count (6, 16, and 32 as one of the first four separate RotEqNet layers, 32 in the additional RotEqNet layer, and 128 in the learn rate pattern) are further unique properties. The use of a 30 approach and 90 100-step time increments was used...

D. . Using a local binary pattern system to identify face fraud

The main contact kindly provided us with the Mat lab code for this model. We made mention to this when, in fact, the system was written in Python using Keras.

Due to the challenging manufacture of this product, testing was done with a small lot size of 20. The space requirements for larger batch sizes were too great for our equipment to manage..

The findings section displays the results column. There were two different features: the ricochet and the frequency of screens (32 tracks in the first unit and 64 tracks after that). In the author's illustration, the decay rate must be learned rather quickly, starting at 0.001 and a half with each generation. The constant were upgraded over the prior versions by reducing the learner rate's velocity vector to 0.01 and boosting the starting rate by 0.8 each epoch.

A few of the several possibilities that were investigated included altering the data set, offering a quick connection from the convoluted module to the thick layers, placing the convective mod well after the gateway tier module, and including extra convolutions. The Particle swarm optimization was utilized in place of stochastic descent, however it did not result in any benefits..

All testing was done on a 64-bit Linux system with two Nvidia GPUs (one Titan Xp: 12GB RAM, 3840 CUDA cores; and one Titan V: 12GB RAM, 5120 CUDA cores). All models were executed on a single GPU. From four to seventy-two hours were spent on aerobic exercises.

Each of the four techniques was created in Python and works only with the parameters specified by the composers (such as learning agreed scope, weight decomposition, velocity, profiler, interlayer, and the number of filtering systems) on each of the three learning curves described in Section 3. The outcomes of these experiments, designated as "original" hyper-parameters, are shown in Table 3. After model parameter values were changed as described in Table 3, the best solutions found are also provided there..

The test data is quite unequal, with an overall percentage of 66% healthy cells, hence the F1-score is most suited for evaluating results. The LBCNN
[3] and DENSENet [5] algorithms can be claimed to be more competent than modern CNNs, VGG, and ResNet [8], which haven't been particularly tuned for skin categorization....

We have our concerns that this is the case because all of our designs had excellent prediction results (LBCNN: 96%, LPBSum: 94%, RotEqNet: 83%) but had much lesser accuracy on the testing set. This problem may be resolved by labeling each cell separately and by gathering more samples, which would give the instructional numbers a greater range of variation among cups and prevent certain glass qualities from being associated with a particular label.

It is important to note that 100% clarity at the cell level is not required as long as proper limitations for evaluating a case can always be established. When a smaller proportion of the germs are found to be cancerous, the patient may be regarded to be in good health, and the collection may be marked for human inspection because when the percentage is higher, the cost of evaluating samples may be greatly reduced by this automated from previously..

V. SIMULATION AND RESULTS

The test set was used to assess the object detection models, and the outcomes are shown in Table 3. The accuracy and loss of the densenet-based object detection models were 84 percent and 0.4, respectively.

This section displays the findings from epochs 1, 10, and 20. As we kept increasing the amount of epochs in the systems, the accuracy and other aspects continued to become better. The dataset illustrations, accuracy and loss curves, and confusion matrices for the 1, 10, and 20 epochs, respectively, are shown in the figures from figure 8 to figure 16..

A. Epoch =1

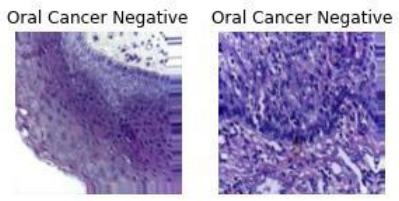


Figure 8 Dataset showing illustration of results in oral cancer negative and positive

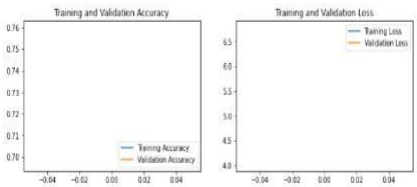


Figure 9 Training loss and training accuracy when epoch =1

63/63 [==============================] No. of steps in Epoch

371s 6s/step - loss: 4.0257 - accuracy: 0.7600 Accuracy on the Test Set = 76.00 % Model Saved!

True: [0 0 0 ... 1 1 1]

Predicted: [0 0 0 ... 0 1 1]

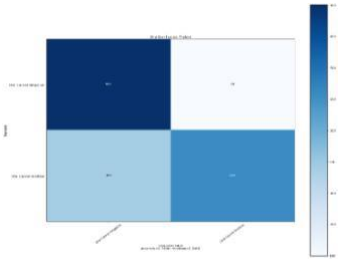
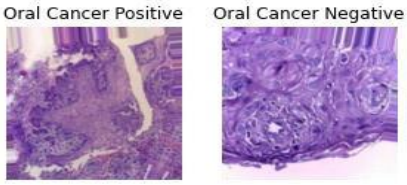


Figure 10 Confusion matrix when epoch =1

Epoch =10



i

Figure 11 Dataset in epoch 10 showing Oral cancer positive and oral cancer negative

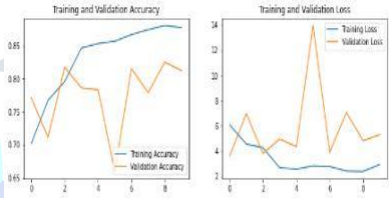


Figure 12 Loss and accuracy recorded at epoch =10

63/63 [==============================] No. of steps in Epoch

378s 6s/step - loss: 5.3044 - accuracy: 0.8125 Accuracy on the Test Set = 81.25 % Model Saved!

True: [0 0 0 ... 1 1 1]

Predicted: [1 0 1 ... 1 1 1]

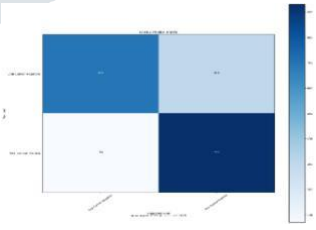


Figure 13 Confusion matrix at epoch

C. Epoch=20

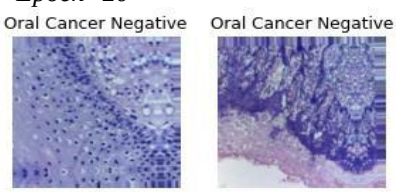


Figure 14 Dataset showing illustration of results in oral cancer negative and positive at epoch 20

Figure 15 Loss and accuracy at epoch 20

63/63 [==============================] No. of steps in Epoch

379s 6s/step - loss: 5.0107 - accuracy: 0.8320 Accuracy on the

Test Set = 83.20 % Model Saved!

True: [0 0 0 ... 1 1 1]

Predicted: [0 0 1 ... 0 1 1]

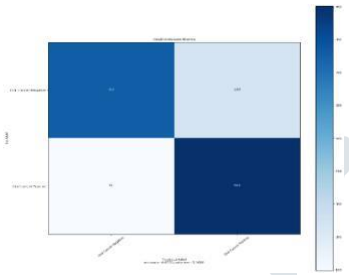


Figure 16 Confusion matrix at epoch 20

Table 3 result summary

Epoch	Accuracy	Loss
Epoch 1	76%	4.02
Epoch 10	81%	0.6
Epoch 20	84%	0.4

This shows that the DenseNet-121 models have potential for the classification and detection of cancer in oral photographs

VI. CONCLUSION

Three state-of-the-art texture-based CNN architectures were implemented, tuned, and evaluated on pictures of oral cancer cells. VGG and ResNet, two general-purpose CNNs, generated results that were compared to those of.

The two algorithms on DENSENET perform better than cutting-edge technology. All-purpose CNNs demonstrate the effectiveness of DENSENETs for this type of photo categorization task as measured by F1 scores. The vector field method RotEqNet produces less promising findings, indicating that it might not be as efficient at categorizing single-cell images. The best performance was provided by the LBCNN model proposed by [13], with an F1 dependability of 81.03% and a result of 84.85%..

Experts will undoubtedly find CNN's robust photo classification capabilities to be a very beneficial method for detecting cancer, reducing their labor, and maybe even paving the road for a nationwide mouth cancer screening program that will lead to a cure. The results of this study demonstrate that dense networks can do better in this task than traditional CNNs..

REFERENCES

[1]Kaiming He, XiangyuZhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[2]FelixJuefei Xu, Vishnu Naresh Boddeti, and MariosSavvides. Local binary convolutional neural networks (lbcnn). https://github.com/juefeix/ lbcnn.torch, 2016.

[3]Felix Juefei-Xu, Vishnu Naresh Boddeti, and MariosSavvides. Local binary convolutional neural networks. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, volume 1. IEEE, 2017.

[4]Gil Levi and Tal Hassner. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In Proceedings of the 2015 ACM International Conference on Multimodal Interaction, pages 503–510. ACM, 2015.

[5]Lei Li, Xiaoyi Feng, Zhaoqiang Xia, Xiaoyue Jiang, and Abdenour Hadid. Face spoofing detection with local binary pattern network. Journal of Visual Communication and Image Representation, 54:182–192, 2018.

[6]LiLiu, Paul Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietik¨ainen. Local binary features for texture classification: taxonomy and experimental study. Pattern Recognition, 62:135–160, 2017.

[7]LiLiu, Songyang Lao, Paul W Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietik¨ainen. Median robust extended local binary pattern for texture classification. IEEE Transactions on Image Processing, 25(3):1368– 1381, 2016.

[8]Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In ICCV, pages 5058–5067, 2017.

[9]Diego Marcos, Michele Volpi, and Devis Tuia. Learning rotation invariant convolutional filters for texture classification. In Pattern Recognition (ICPR), 2016 23rd International Conference on, pages 2012–2017. IEEE, 2016.

[10]Timo Ojala, Matti Pietik¨ainen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1):51–59, 2016.

[11]TimoOjala, Matti Pietikainen, and Topi Maenpaa. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, 2002.

[12]PeterSasieni,Alejandra Castanon, and Jack Cuzick. Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. BMJ, 339:b2968, 2009.

[13]Olivier Simon, Rabi Yacoub, Sanjay Jain, John E Tomaszewski, and Pinaki Sarder. Multi-radial DENSENET features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. Scientific Reports, 8(1):2032, 2018.