





Research Article

An Automatic Mass Screening System for Cervical Cancer Detection Based on Convolutional Neural Network

Aziz-ur -Rehman ¹, Nabeel Ali ², Imtiaz.A. Taj ², Muhammad Sajid ³,
and Khasan S. Karimov^{1,4}

¹Department of Electrical Engineering, GIK Institute of Engineering Sciences and Technology, Topi 23640, District Swabi, Khyber Pakhtunkhwa, Pakistan

²Department of Electrical Engineering, Capital University of Science and Technology, Islamabad Expressway, Kahuta Road, Zone-V, Islamabad, Pakistan

³Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur 10250, Pakistan

⁴Centre for Innovative and New Technologies of Academy of Sciences of the Republic of Tajikistan, Rudaki Ave., 33, Dushanbe 734015, Tajikistan

Correspondence should be addressed to Aziz-ur -Rehman; azizur80@hotmail.com

Received 19 May 2020; Revised 30 September 2020; Accepted 8 October 2020; Published 28 October 2020

Academic Editor: Jose Merodio

Copyright © 2020 Aziz-ur -Rehman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cervical cancer is the fourth most common type of cancer and is also a leading cause of mortality among women across the world. Various types of screening tests are used for its diagnosis, but the most popular one is the Papanicolaou smear test, in which cell cytology is carried out. It is a reliable tool for early identification of cervical cancer, but there is always a chance of misdiagnosis because of possible errors in human observations. In this paper, an auto-assisted cervical cancer screening system is proposed that uses a convolutional neural network trained on Cervical Cells database. The training of the network is accomplished through transfer learning, whereby initializing weights are obtained from the training on ImageNet dataset. After fine-tuning the network on the Cervical Cells database, the feature vector is extracted from the last fully connected layer of convolutional neural network. For final classification/screening of the cell samples, three different classifiers are proposed including Softmax regression (SR), Support vector machine (SVM), and GentleBoost ensemble of decision trees (GEDT). The performance of the proposed screening system is evaluated for two different testing protocols, namely, 2-class problem and 7-class problem, on the Herlev database. Classification accuracies of SR, SVM, and GEDT for the 2-class problem are found to be 98.8%, 99.5%, and 99.6%, respectively, while for the 7-class problem, they are 97.21%, 98.12%, and 98.85%, respectively. These results show that the proposed system provides better performance than its previous counterparts under various testing conditions.

1. Introduction

Cervical cancer is the leading cause of cancer-related deaths in females. It arises from the cervix, i.e., the lower and narrow end of the uterus, as shown in Figure 1. It starts due to abnormal growth of cells that have the ability to spread into other parts of the body. Human Papilloma Virus (HPV) infection is the major risk factor for cervical cancer. There are no symptoms in the beginning of the disease, while with the passage of time symptoms may include abnormal vaginal bleeding, pelvic pain, and pain during sexual intercourse. It

can be diagnosed earlier through regular medical check-ups [1].

There are many diagnostic tests for cervical cancer identification. Papanicolaou (PAP) smear is the most commonly used test for cervical cancer screening worldwide. In the conventional PAP smear procedure [2], a speculum is inserted into vagina to widen the walls so that vaginal smear can be viewed. Several weeks are required to prepare for the final results of the PAP smear test. The process is time-consuming and laborious. It requires microscopic examination of hundreds of thousands of cells for the diagnosis of

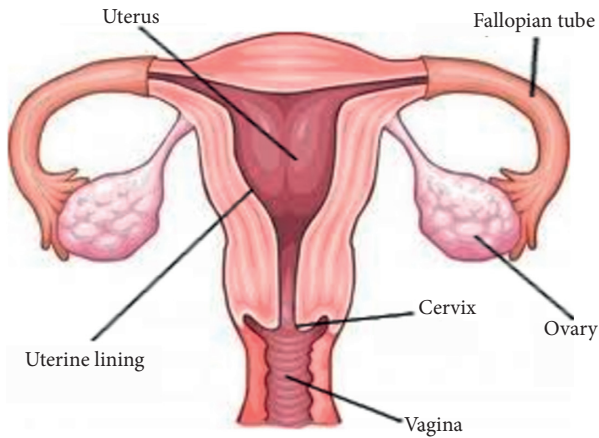


FIGURE 1: Female reproductive system. Initially, the cervical cancer starts from the cervix, the lower end of uterus.

precancerous and cancerous cells. In every 10 to 15 positive cases, there is a chance of one case to be missed in conventional screening [3].

The rate of the incidence of cervical cancer is lesser in the USA and other parts of developed countries because of early detection and better screening methods [4]. Its rate of occurrence has been dropped by 80% since the screening systems are introduced in some Nordic countries. In Sweden [5], it is dropped by 65% during the last four decades and the occurrence of cervical cancer and mortality figures are stable over the last decade. However, improved screening systems are still unavailable in underdeveloped countries, partly due to the complexity and tedious nature of manual screening of abnormal cells from a cervical cytology specimen [6, 7]. While auto-assisted mass screening techniques can boost efficiency, they are not accurate enough to be used as a primary tool for cervical screening [8].

During the past few years, extensive research has been carried out for the development of computer-assisted automated reading systems based on cell image analysis [7, 9, 10]. The manual screening process is normally initiated with the collection of cervical cell samples from the uterine cervix and their placement on a glass slide. After visual inspection under a microscope, these are classified into different categories. The shape, size, texture, ratio of nucleus, and cytoplasm are the main characteristics for the classification task. Hence, for an automated system, the first step may include segmentation of images of cell samples to extract regions of interest, containing single cells with nucleus and cytoplasm, from the noncell regions. This initial segmentation is then followed by separation of main cell components including the nucleus and cytoplasm and extraction of their shape/textural features. However, the separation of main components and shape feature extraction is not an integral part of an automated screening system, as proposed schemes in the literature include both options, i.e., with geometrical feature extraction and without prior extraction.

For a system that includes prior feature extraction, accurate segmentation of nucleus from cytoplasm in cervical cell images is a difficult task and is prone to error, thus

limiting the success of overall system. The presence of large irregular shapes, appearance dissimilarities, and cell clusters between malignant and benign cell nucleus is the major problem in accurately segmenting the cytoplasm and nucleus. Various segmentation algorithms have been proposed by researchers to segment out cell components. An iterative algorithm for assigning pixels based on a statistical criterion function was proposed in [11] to separate the nucleus, cytoplasm, and background. In another study [12, 13], Gabor filters were applied for exploiting textural variation of the cervical cells to segment out regions of interest. Fuzzy C-means clustering was used in [14, 15] to segment the single cell images into nucleus, cytoplasm, and background. However, if the overlapping cells are taken into account, the classification accuracy is decreased significantly. Therefore, a majority of the presented segmentation approaches [16–20], [11, 12, 14] are effective in terms of their performance for single and clear cervical cell images only, but in the case of overlapping cells or other shape changes, they lack the performance accuracy.

To overcome this dependency on segmentation, many techniques have been proposed during the past few years, which do not include prior segmentation and directly classify the unsegmented cell images. A pixel-level classification method is proposed in [21] to classify normal and abnormal cells without prior segmentation using block-wise feature selection and extraction techniques. However, the validation accuracy of the proposed algorithm is not up to the mark. In [22], block image processing was proposed that includes cropping arbitrary image blocks prior to feature extraction, and the cropped blocks are then classified using SVM. However, in their approach, arbitrary cropping could potentially separate a full cell into distinct patches.

Recently, feature representation in image classification problems based on deep learning methods has become more popular [23]. In particular, convolutional neural networks (ConvNets) [24] have achieved unprecedented results in the 2012 ImageNet Large Scale Visual Recognition Challenge, which consisted of classifying natural images in the ImageNet dataset into 1000 fine-grained categories [25]. Besides, they have drastically increased the accuracy in the field of medical imaging [26, 27], specifically classification of lung diseases and lymph nodes in CT images [28, 29], and detecting cervical intraepithelial neoplasia based on cervigram images [30] or multimodal data [31]. ConvNets have also shown superior performance in the classification of cell images for diagnosis of pleural cancer [32].

However, large datasets are essential to achieve high performance and to overcome the problem of overfitting with ConvNets [33]. This is a major limitation in applying ConvNets to the cervical cell classification problem as in the case of cervical cells, and a limited number of annotated datasets are available. For instance, the Herlev dataset [34] only contains 917 cervical cells with 675 abnormal and 242 normal cells that are insufficient for ConvNets. To overcome this limitation, recently, image data augmentation techniques have been proposed to virtually increase the size of training datasets and reduce the problem of overfitting [25]. Data augmentation can be achieved by linear transformation

of the data such as mirroring, scaling, translations, rotation, and color shifting unless the information of the object in the image is intact. Transfer learning [21, 22, 35–39] is another solution to overcome data overfitting. In transfer learning, a ConvNet is first trained on large-scale natural image datasets and then can be fine-tuned to the desired dataset which is limited in the size.

In this paper, an automatic screening system is proposed to classify malignant and benign cell images without prior segmentation using ConvNets. Due to limited size of Herlev datasets, transfer learning is used to initialize the weights and then fine tune on the dataset. The feature vector at a fully connected layer is extracted after fine-tuning and passed to various classifiers. To show the efficacy of the proposed approach, its performance is evaluated on the Herlev dataset for 2-class and 7-class problems. Malignant and benign cells are considered in the 2-class problem, while in the 7-class problem, all seven categories of the cervical cells have been explored. In short, the research contributions of the presented work are summarized as follows:

- (1) Our work is aimed at developing tool for automatic classification of cervical cells using the convolutional neural network. Unlike previous methods, it does not require prior segmentation and hand-crafted features. This method automatically extracts hierarchical features embedded in the cell image for the classification task.
- (2) A data augmentation technique has been considered to avoid overfitting. The rate of overfitting has been reduced as the data augmentation strategy is applied to train our network. This approach is fruitful for our network to learn the most discriminative features of cervical cells and thus achieve superior classification results.
- (3) Transfer learning is also explored for pretraining, and initial weights are reassigned to another network for fine-tuning on cervical cell images. Training from scratch requires a large amount of labeled data which is extremely difficult in medical diagnosis. Moreover, the designing and adjustment of the hyper-parameters are the challenging tasks with reference to overfitting and other issues. Transfer learning is the easiest way to overcome such problems.
- (4) We also conduct extensive malignant and benign cell assessment experiment on the Herlev dataset. Our results clearly demonstrate the effectiveness of the proposed convolutional neural architecture. The experimental results are compared with recently proposed methods, and our approach provides superior performance as compared with existing systems for cervical cells classification.

The paper is organized as follows: the proposed methodology is presented in Section 2; experiments and results are given in Section 3; result-related discussion is presented in Section 4; and conclusion and future work are summarized in Section 5.

2. Proposed Methodology

The proposed automatic mass screening system for cervical cancer detection using ConvNets is shown in Figure 2. There are four steps: (1) data collection, (2) preprocessing, (3) feature learning, and (4) classification of cervical cells. These steps are explained in the following sections.

2.1. Data Collection. The publicly available Herlev Pap smear dataset is used for the training and testing purpose. It contains 917 single cervical cell images with ground truth classification and segmentation. The cells are categorized into seven different classes. These seven classes are diagnosed by doctors and cytologists to increase the reliability of the diagnosis. Furthermore, these seven classes are broadly categorized into two groups, i.e., malignant and benign. The first class to third class is normal or benign, while fourth to seventh class is abnormal or malignant. The class's distribution is shown in Table 1.

Normal and abnormal cell images are shown in Figure 3. It can be seen that the size of the nucleus in malignant or abnormal cells is larger than that of the normal cells. The difficult task from classification perspective is that the normal columnar cells have nucleus size quite similar to that of severe nucleus, and also chromatin distribution is same.

2.2. Preprocessing. Herlev dataset consists of images that contain multiple cells in a single image. The data preprocessing phase includes image patch extraction from the original cervical cell images and augmentation of data for training ConvNet 2.

2.2.1. Image Patch Extraction. The proposed approach, like previous patch-based classification methods, does not directly operate on original images present in the Herlev dataset that contains multiple cells at a time [40–43]. Image patches, each containing single cell, are first extracted. In order to extract the individual cell, presegmentation of cytoplasm is required [44]. The nuclei are first detected and then image patches of size $M \times M$, and each centered on a nucleus is generated that embed not only the size and scale information of the nucleus but also the textural information, of the cytoplasm surrounding the nucleus. Scale and size of the nucleus is a very important discriminative feature between malignant and benign cells.

2.2.2. Data Augmentation. An image data augmentation technique is used to virtually increase the size of training dataset and reduces overfitting [25]. As the cervical cells are invariant to rotations, they can be rotated from 0 – 360 degree with a step angle θ . In the data augmentation process, $N_r = 10$ rotations with $\theta = 36$ degree, translations in the horizontal direction, $N_{th} = 15$ translations up to 15 pixels for each normal cells, while in vertical direction, $N_{tv} = 8$ translations up to 15 pixels for each abnormal cells are performed. Hence, we generate 300 image patches from a

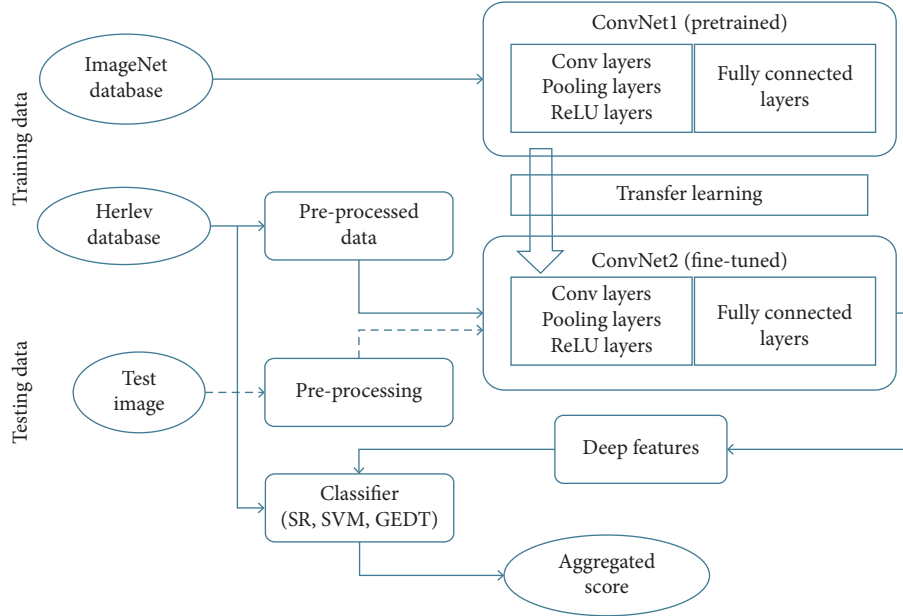


FIGURE 2: The process of cervical cancer detection system using ConvNets through transfer learning.

TABLE 1: Key statistics of Herlev dataset [36].

Type	Class	Cell description	Numbers (cells)
Normal	1	Superficial squamous epithelial	74
Normal	2	Intermediate squamous epithelial	70
Normal	3	Columnar epithelial	98
Abnormal	4	Mild squamous nonkeratinizing dysplasia	182
Abnormal	5	Moderate squamous nonkeratinizing dysplasia	146
Abnormal	6	Severe squamous nonkeratinizing dysplasia	197
Abnormal	7	Squamous cell carcinoma in situ intermediate	150

single normal cell and 160 image patches from each abnormal cell. This transformation yields relative normal distribution, as the numbers of samples of abnormal cell images are as large as compared to that of normal cell images. The size of the generated image patch is set to $m = 128$ pixels to cover the cytoplasm region. These patches are then upsampled to a size $256 \times 256 \times 3$ using bilinear interpolation. These upsampled image patches, as shown in Figure 4, are used in ConvNet 2 for initiating layer transfer and training [28].

The malignant cells in the Herlev dataset are 3 times more than the benign cells. Therefore, it is natural that the classifier tends to be more biased towards the majority class, i.e., the malignant cells. The unfair distribution of data is commonly solved by normalization of data prior to classification, whereby the ratio of positive and negative samples of data is evenly distributed [45]. This normalization process improves not only the convergence rate of training of the ConvNets but also the classification accuracy [25]. In the proposed approach, the training dataset is made balanced by unequal augmentation of benign and malignant cells, in which a higher proportion of benign training samples are generated as compared to malignant training samples.

2.3. Feature Learning. The ConvNets can learn to discriminate features automatically for an underlying task. In this work, a typical deep model is used consisting of 2 ConvNets, named ConvNet 1 and ConvNet 2. At first, the base network ConvNet 1 is pretrained on ImageNet database that consists of over 15 million labeled high-resolution images, belonging to roughly 22,000 categories [46]. The images were collected from the web and labeled through human judgment using Amazon's Mechanical Turk crowdsourcing tool. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images [25]. ConvNet 1 contains five convolutional (conv) layers denoted as conv1–conv5, followed by three pooling (pool) layers denoted as (pool1), (pool2), and (pool5), and there are three fully connected (fc) layers as (fc6), (fc7), and (fc). All these layers are transferred to ConvNet 2, which is the network used for feature extraction, setting its initial parameters. This new network is then fine-tuned on the single cervical cell images of the Herlev database. This procedure is shown in Figure 5.

As described earlier, (conv) and (pool) layers are transferred from ConvNet 1 at the same locations to ConvNet 2. Both ConvNet 1 and ConvNet 2 share the same structure from (conv1) to (pool5) layers. However, the fully

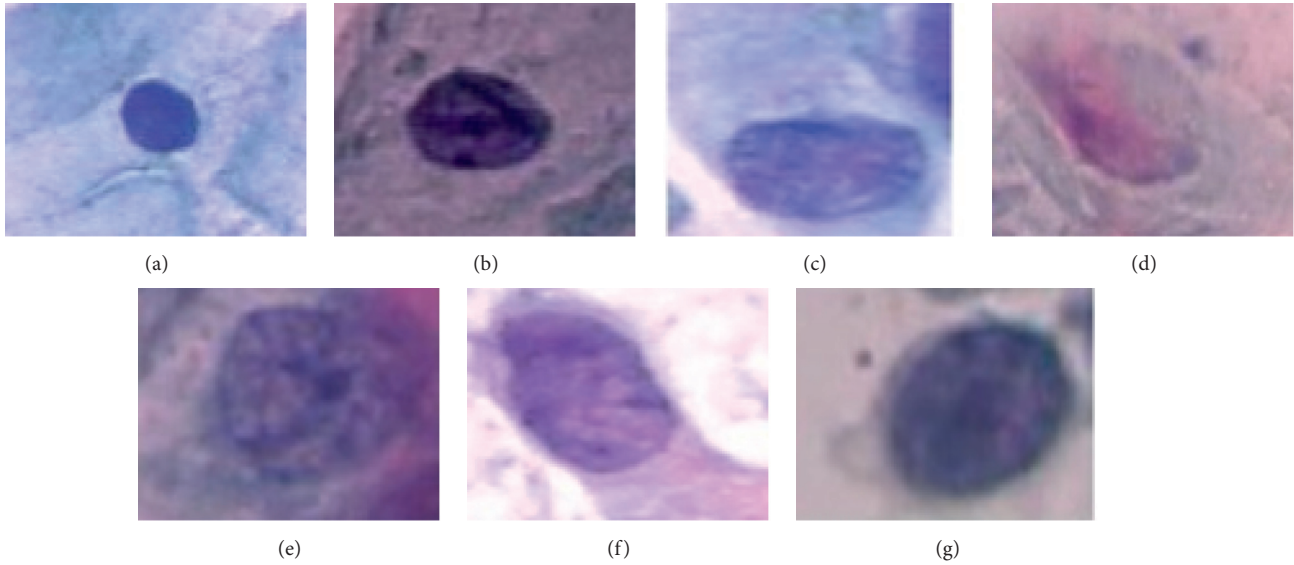


FIGURE 3: Examples of (a–c) normal cells and (d–g) abnormal cells: (a) superficial; (b) intermediate; (c) columnar; (d) mild; (e) moderate; (f) severe; (g) carcinoma.

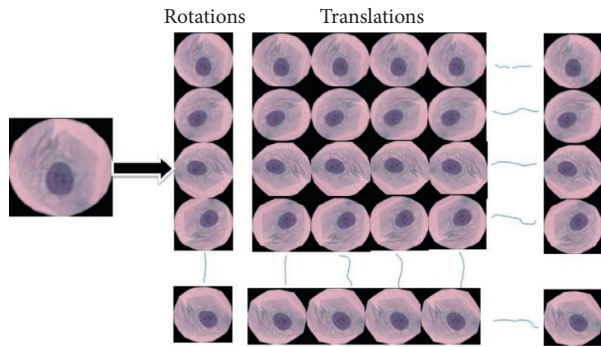


FIGURE 4: Data augmentation process through rotations and translations of a cervical cell.

connected layers are modified in ConvNet 2 because the number of output classes is different as compared to ConvNet 1. Numbers of neurons in ConvNet 2 are $4096 - 7$ and $4096 - 4096 - 2$ in the case of 7-class problem and 2-class problem, respectively. Fully connected layers of ConvNet 2 are initialized with values from random Gaussian distributions. Local response normalization for (conv1) and (conv2) is set according to the parameters in [25]. Hidden layers are used with rectified linear units' activation function. There are three fully connected layers in the proposed network, i.e., (fc6), (fc7), and (fc). The feature vector for the final classification task is selected from the (fc7) layer which is the last layer before the output layer. The main reason to select feature vector from (fc7) is that it contains more specific and abstract details of the images. The dimension of feature vector extracted from (fc7) is the number of training samples $X_t \times 4096$. The configuration of ConvNet 2 is listed in Table 2.

2.4. Classification. Deep features are extracted from the outer layer of ConvNet 2 for cervical cells classification task.

The classification score is then calculated using three different classifiers including SR, SVM, and GEDT. The details of three classifiers are also presented as follows.

2.4.1. Softmax Regression (SR). For the multiclass dataset, SR is used for classification of unknown samples that are first preprocessed according to the described approach. Unlike ConvNets, SR uses cross entropy function for the classification. The sigmoid function is replaced by softmax function. Mathematically, it is represented by the following equation:

$$P(y = jz^{(i)}) = \phi_{\text{softmax}}(z^{(i)}) = \frac{e^{z^{(i)}}}{\sum_{j=0}^k e^{z_k^{(i)}}}, \quad (1)$$

where we define the network input z as

$$z = w_0x_0 + w_1x_1 + \dots + w_nx_n = \sum_{i=0}^n w_ix_i = w^T x, \quad (2)$$

where w is a weight vector, x is the feature vector of training sample, and w_0 is the bias. This softmax function computes the probability score that training sample x_i belongs to class j given the network z . The probability score is generated at the softmax layer of ConvNet 2, next to fully connected layer. Cross entropy function is used for the classification at the final layer of the ConvNet 2. Softmax layer of ConvNet 2 is shown in Figure 6.

2.4.2. Support Vector Machine (SVM). SVM is a supervised learning model that uses an optimization method to identify support vectors S_i , weights α_i , and bias b . The classification is being considered to classify vectors x , according to the following equation:

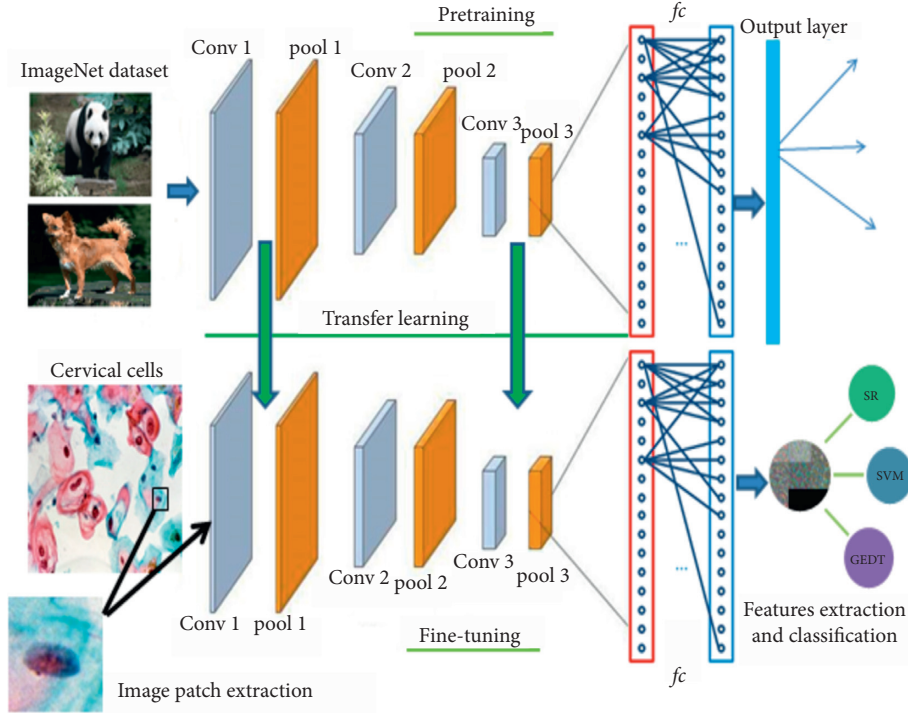


FIGURE 5: Procedure of feature learning and classification using convolutional neural networks.

TABLE 2: Configuration of ConvNet 2.

	Filter size	Channel	Stride	Padding
Input	—	3	—	—
(conv1)	11 × 11	96	4	—
(pool1)	3 × 3	96	2	—
(conv2)	5 × 5	256	1	2
(pool2)	3 × 3	256	2	—
(conv3)	3 × 3	384	1	1
(conv4)	3 × 3	384	1	1
(pool5)	3 × 3	256	2	—
(fc6)	—	4096	—	—
(fc7)	—	4096	—	—
(fc)	—	7 (7 classes) 2 (2 classes)	—	—

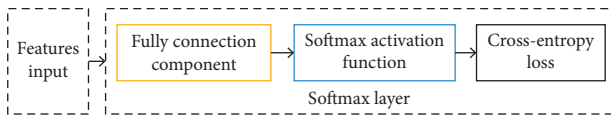


FIGURE 6: Softmax layer of ConvNet 2.

$$c = \sum_i \alpha_i k(s_i, x) + b, \quad (3)$$

where k is the kernel function depending on the model assumed for decision boundary. In case of a linear kernel, k is a dot product. If $c \geq 0$, then x is classified as a member of the first group and otherwise the second group. Error correcting code classifier is trained using support vector machine. The batch size is set to 256. The training set is applied to the

classifier along with deep hierarchical feature vector using ConvNet 2. Validation data are used in SVM to calculate validation accuracy.

2.4.3. GentleBoost Ensemble of Decision Tree (GEDT). Regression trees or GEDT are used to predict the response of the data. The classification decisions are made when the query sample follows the path from the initial or root node to the end or leaf node.

In GEDT, an ensemble of trees is used that is based on majority voting. It is trained on the training data, numbers of trees are set to 100, and batch size is set to 256. Validation accuracy of the Herlev dataset is evaluated using validation data.

2.4.4. Aggregated Score. Evaluation of the cervical cell classification task is done using 5-fold cross validation on the Herlev dataset for both 2-class and 7-class problems. The performance metrics used for evaluation include accuracy, F1 score, area under the curve, specificity, and sensitivity. Finally, the count of correct classification score is obtained for each cell from all the categories in the Herlev dataset.

3. Experiments and Results

3.1. Experimental Protocol. In the training stage, the conv and pool layers of Alexnet, i.e., ConvNet 1, as shown in Figure 5, are used as initial layers for the ConvNet 2. Random weights are initialized to fc . In order to train ConvNet 2, a patch of size 227×227 is cropped randomly from each augmented image to make training/test images

compatible to the input nodes of the network. To achieve zero-center normalization, a mean image over the dataset is subtracted. Stochastic gradient descent (SGD) is used for training ConvNet 2 using 30 epochs. Small batches of image patches are fed to ConvNet 2, and validation accuracy of batches is evaluated. The size of mini batch is set to 256. Initial learning rate for convolutional and pooling layers is set to 0.0001, which is set to decrease with a factor of 10 after every 10 epochs. L2 regularization and momentum can be tuned to reduce overfitting and speed up learning process of the ConvNet 2 [25]. L2 regularization or weight decay and momentum are empirically set to 0.0005 and 0.9, respectively. Finally, the network is trained using a randomly selected subset of epochs and validated for its accuracy. The model having a minimum validation error is used for classification application.

In order to test the system against an unseen image, multiple cropped patches of test images, each having single cell, are generated from the original images containing multiple cells. Abnormal score of each crop is generated by the classifier. The abnormal scores of all ($N_{\text{test}} \times N_{\text{crop}}$) patches of the test image are aggregated to generate the final score [47]. Patches of test image $N_{\text{test}} = 300$ (10 rotations \times 30 translations) are generated same as for training images. Furthermore, ten cropped images (N_{test}) (four corner, center of cell, i.e., nucleus portion, and their mirrored images) are generated from each of test patch. These ($N_{\text{test}} \times N_{\text{crop}}$) image patches are input to the classifier. The prediction score of the classifiers (SR, SVM, and GEDT) is then aggregated to calculate the final score, as shown in Figure 7.

3.2. Experimental Results and Evaluation

3.2.1. ConvNet 2 Learning Results. ConvNet 2 is fine-tuned on the Herlev dataset for 2-class and 7-class problems using 30 epochs. It is observed that, after 10 epochs, the validation accuracy reaches its maximum value, i.e., 0.9935 for the 2-class problem and 0.8745 for the 7-class problem. Figure 8 illustrates a fine-tuning process of ConvNet 2 during 30 training epochs.

These results are improved by considering various classifiers. GEDT provides better performance with reference to both classes because it exploits the randomness of data more efficiently as compared to other classifiers. Table 3 shows the comparison of SR, SVM, and GEDT.

The structure of the layers of network is also being observed after passing a test image to the fine-tuned network. Features learned at the first layer, i.e., (conv1) are more generic, as shown in Figure 9.

It can be seen that these learned filters contain gradients of different frequencies, blobs, and orientations of colors. As we go deeper in convolutional layers (conv1 – conv5), the features become more prominent and provide more information. Figure 10 shows the feature learning results in (conv2 – conv5) for a test cervical cell image.

The strongest activation is also shown in Figure 11 at the pool layer. The white pixels show strong positive activation, while black pixels provide strong negative activation and

gray does not activate strongly. It is also observed that the strongest activation initiates negatively on right edges and positively on left edges.

The feature set at fully connected layer is also explored and it is observed that features are more abstract as compared to the previous layers, as shown in Figure 12. Fully connected layer provides features learned for 7 classes.

3.2.2. ConvNet 2 Testing Results. It is a common knowledge that a single evaluation metric is not appropriate to evaluate the performance of a given algorithm due to the presence of some imbalanced classes in the dataset or a large number of training labels [48]. Therefore, the performance of the deep model is reported in terms of four distinct metrics including accuracy, sensitivity, specificity, and F1 score, as proposed in the previous studies [49]. These performance parameters are calculated using the following equations:

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{TP + FP + TN + FN}, \\ \text{sensitivity} &= \frac{TP}{TP + FN}, \\ \text{specificity} &= \frac{TN}{TN + FP}, \\ F1 &= 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \end{aligned} \quad (4)$$

where the precision and recall are expressed as

$$\begin{aligned} \text{precision} &= \frac{TP}{TP + FP}, \\ \text{recall} &= \frac{TP}{TP + FN}. \end{aligned} \quad (5)$$

In the above equations, true positive (TP) is defined as the number of malignant cell images classified as malignant and true negative (TN) is the number of benign cell images classified as benign. False positive (FP) is the number of benign cell images identified as malignant cell images, and false negative (FN) is the number of malignant cell images classified as benign.

In ConvNet 2, a test set of cervical images using the multiple crop testing scheme is considered with three classifiers, i.e., SR, SVM, and GEDT. It can be seen that GEDT again outperforms the classification accuracy of SR and SVM in test results also. The results are presented in Table 4.

By analyzing the class-wise accuracy, one can observe that the proposed method can predict the cervical cell images well. The classification accuracy of each of the seven cell categories is calculated by feeding all the images as test to the classifiers. It is observed that GEDT shows superior performance on class 1, class 2, class 4, and class 5 because of its ability to eliminate irrelevant features and to extract decision rules from decision trees efficiently. The performance slightly deteriorates for class 3 and class 6 because their features are very close to each other causing confusion. The

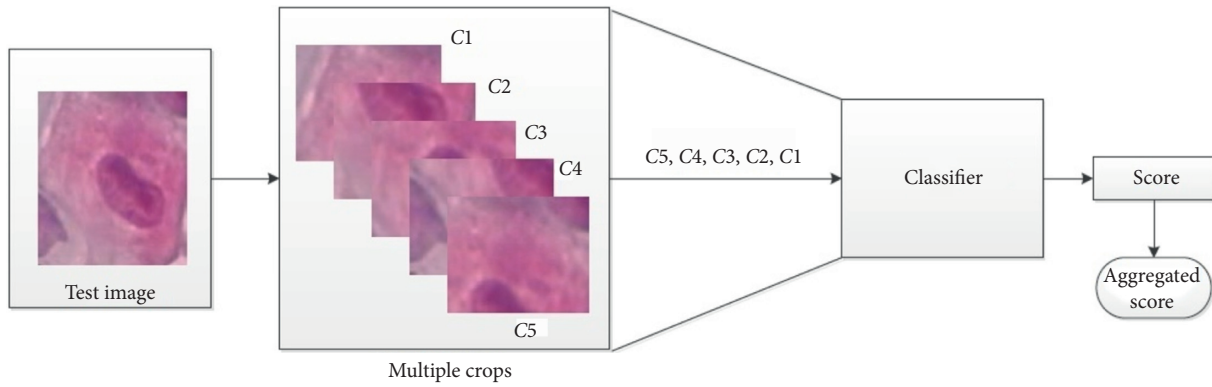


FIGURE 7: Multiple crop testing scheme.

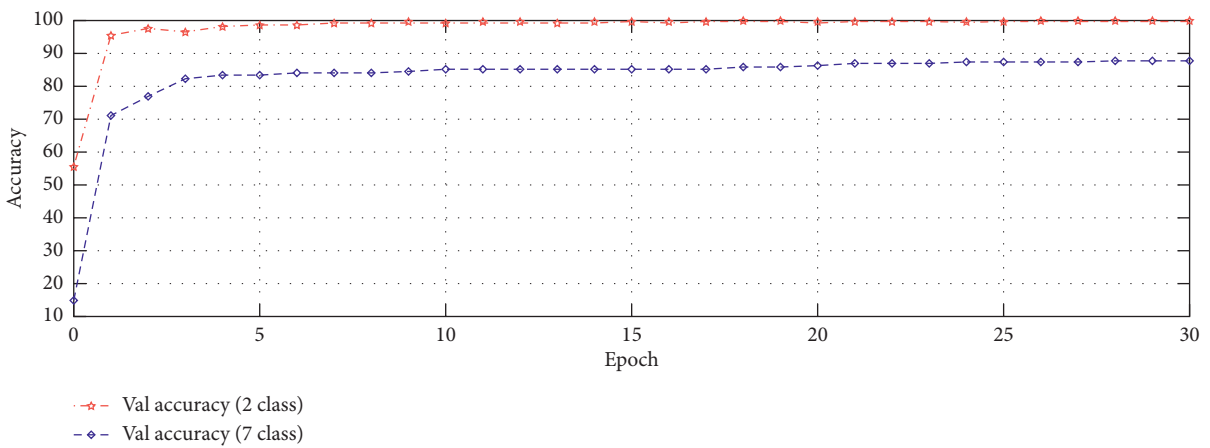


FIGURE 8: The fine-tuning process of ConvNet 2 during 30 epochs for 2-class and 7-class.

TABLE 3: The validation accuracy comparison of three classifiers.

Classes	Validation accuracy		
	SR	SVM	GEDT
2-class problem	99.35	99.8	100
7-class problem	87.45	96.20	99.27

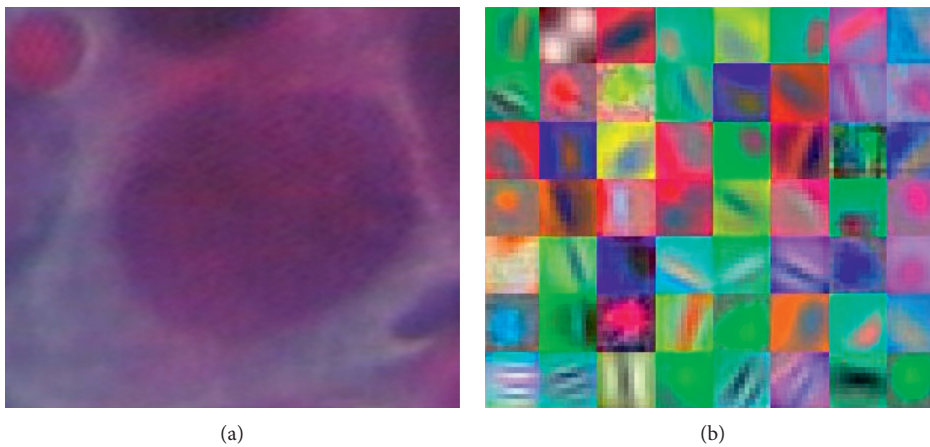


FIGURE 9: (a) Test image; (b) feature learned at (conv1).

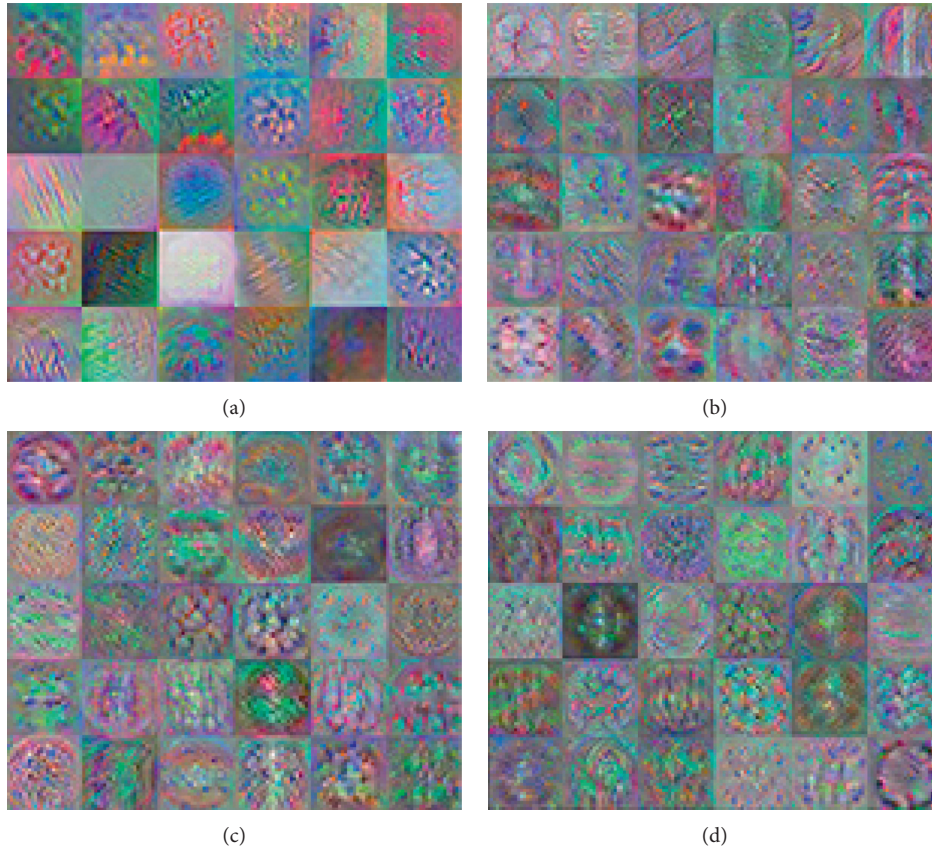


FIGURE 10: Features maps: (a)(conv2); (b)(conv3); (c)(conv4); (d) Conv5.

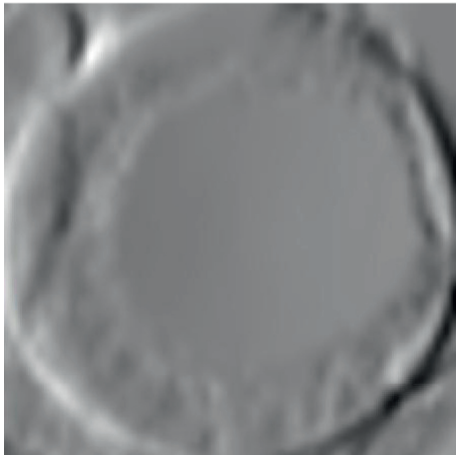


FIGURE 11: Strongest activation at the deepest (pool) layer.

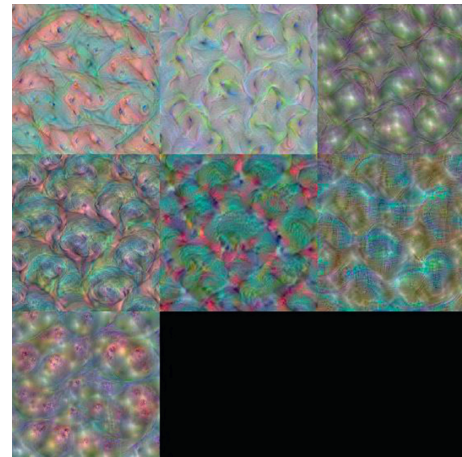


FIGURE 12: Fully connected layer feature maps for 7 classes.

classification accuracy of class 3 and class 6 is 97.50 % and 97.79 %, respectively. Classification accuracy for class 7 is 99.20 %. The average accuracy achieved by GEDT for underlying task is 99.21%. These results are illustrated in Figure 13.

The evaluation parameters of the classification performance, i.e., accuracy, F1 score, area under the curve, specificity, and sensitivity of the trained ConvNet 2 are displayed in Tables 5 and 6, where the performance

comparison of proposed work with [13,39] and [50–56] is presented. We have proposed two scenarios with different classifiers, i.e., SVM and GEDT. The mean values of accuracy, F1 score, area under the curve (AUC), sensitivity, and specificity of fine-tuned ConvNet 2 with GEDT classifier are 99.6%, 99.14%, 99.9%, 99.30%, and 99.35%, respectively, for the 2-class problem. These are 98.85%, 98.77%, 99.8%, 98.8%, and 99.74%, respectively, for the 7-class problem.

TABLE 4: Accuracy after classifier trained on deep features.

Classes	Accuracy		
	SR	SVM	GEDT
2-class problem	99.80	99.50	99.60
7-class problem	97.21	98.12	98.85

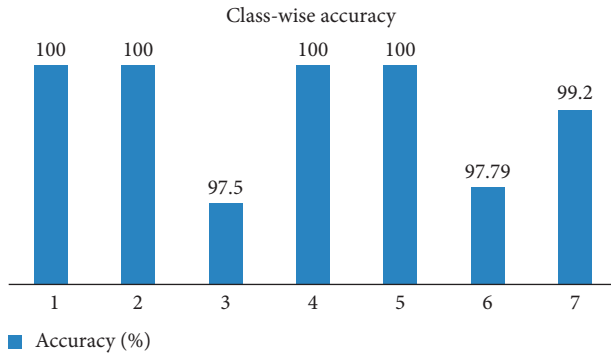


FIGURE 13: Accuracy distributions according to all 7 classes.

The accuracy of our system, i.e., ConvNet with GEDT is 99.6% for 2-class and 98.85% for 7-class compared to 99.5% and 91.2% for [39], respectively. This indicates that the prediction accuracy of our classification model is better than the existing models. Similarly, the sensitivity of 99.38% and 99.30% implies better performance of the proposed method compared to the existing methods in classifying the cervical cell images. Likewise, the values of specificity and accuracy of proposed system for the 2-class problem are better than previous methods in [15, 16, 34, 36, 38, 41, 56].

The images of cervical cells that are correctly classified or misclassified are also analyzed. Figure 14 shows the correctly classified malignant cell images; columns 1 to 4 are mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma, respectively.

Figure 15 shows the result for test cell images that are misclassified (normal misclassified as abnormal and abnormal misclassified as normal).

3.2.3. Computational Complexity. In the training phase, ConvNet 1 is trained on the Corei7 machine with clock speed 2.8 GHz, NVidia 1080Ti GPU, and 8 GB of memory on MATLAB R2017b. The average training time for ConvNet 2, running for 30 epochs, is about 4 hours and 30 minutes for the 2-class and 8 hours and 20 minutes for 7-class problem. In the testing phase, the system takes 8 seconds to classify a test image into normal and abnormal classes. Using multiple crop testing, i.e., ($N_{patches} \times N_{crops} = 3000$) classifications and score aggregation, the average time for the testing of one cell image is around 8 seconds.

4. Results-Related Discussion

The experimental results presented in this study suggest the following key observations:

- (1) Compared with the traditional prior feature extraction schemes, the proposed cervical cell screening system is more effective and robust. This is because ConvNets have been used to encode cervical cancer specific features automatically. In the traditional methods such as [11, 12, 14, 15], cervical cells extraction strategies are hand-crafted which limit the success of overall system. Moreover, in the presence of large irregular shapes, appearance dissimilarities and cervical cell clusters between malignant and benign cell nucleus are the major problems in accurately segmenting the cytoplasm and nucleus. In contrast, this method uses automatic cervical cell features extraction to encode cancerous cell representation and thus achieve superior classification accuracy across a range of cells severity.
- (2) In order to prevent overfitting, a data augmentation technique, suitable for the underlying task of cervical cell grading, has been proposed. The training and validation losses for 30 epochs have been evaluated to analyze the impact of the proposed data augmentation on classification accuracy. It is observed that the rate of overfitting is greatly reduced when the data augmentation strategy is applied to train our classification model. The smaller difference between training and validation losses caused by data augmentation is presented in Figure 16. It indicates that how this approach is fruitful for the classification model to learn the most discriminative features for the desired task. Furthermore, the proposed model works across a variety of cervical cells and preserves the discriminative information during training. While, in the testing stage, a cell image with arbitrary level of severity can be easily classified into the true grading level. Hence, this suggests the efficacy of our method to avoid the classification model from overfitting and shows robustness for classification accuracy against varying nature of cervical cells.
- (3) A multiple crop testing scheme is also used with three classifiers to calculate the accuracy of all individual classes of cervical cell images. The class-wise accuracy displayed in Figure 13 shows, if the cervical cells are more clear, the classification ability of our system is more robust. For example, classification accuracy for class 1, class 2, class 4, and class 5 is the highest, i.e., 100% among all other classes. It is because this type of cells can be identified more effectively by the underlying model. The classification accuracy for class 7 is 99.20%. Conversely, the classification performance slightly degrades for the class 3 (97.50%) and class 6 (97.79%) because their features are very close to each other owing to lesser cervical cells specific discriminative information presented to the model.
- (4) In a general way, a single performance metric can lead to inappropriate classification results due to some imbalanced classes in the dataset or too small or large number of training samples. The literature

TABLE 5: Accuracy comparison of the systems for 2-class and 7-class problems.

Systems	Features	Dataset	Accuracy%
[50]	SVM	149 images of Herlev	98 (2-class)
[51]	k-NNs and ANNs	Herlev	k-NNs: 88; ANNs: 54 (2-class)
[52]	20 features, fuzzy C-means, 2nd order NN	Herlev	98.4 (2-class)
[53]	Decision tree	Local	67.5 (2-class)
[54]	Backpropagation neural networks	Local	95.6 (2-class), 78.06 (7-class)
[55]	20 features, C-means/fuzzy clustering	Herlev	94–96 (2-class), 72–80 (7-class)
[13]	Shape, texture, Gabor, SVM	Local	96 (2-class)
[56]	20 features, GA	Herlev	98 (2-class), 96.9 (7-class)
[39]	CNN-ELM-based system	Herlev	99.5 (2-class), 91.2 (7-class)
Proposed	ConvNet with SVM	Herlev	99.5 (2-class), 98.12 (7-class)
Proposed	ConvNet with GEDT	Herlev	99.6 (2-class), 98.85 (7-class)

TABLE 6: Comparison of the proposed work with existing state-of-the-art using the Herlev dataset.

Systems	Features	Sensitivity (%)	Specificity (%)	Accuracy (%)
[34]	Pap smear benchmarks	98.8	79.3	93.6
[56]	Genetic algorithm-based feature	98.5	92.1	96.8
[16]	PCA with SVM	—	—	96.9 (F1 score)
[41]	Local binary pattern	—	—	96.4 (AUC)
[38]	Ensemble classifier	99.0	89.7	96.5
[36]	CNN	98.2	98.3	98.3
[15]	Enhanced fuzzy C-means	99.28	97.47	98.88
Proposed	ConvNet with SVM	99.38	99.20	99.5
Proposed	ConvNet with GEDT	99.30	99.35	99.6

The performance metrics are evaluated for the 2-class problem.

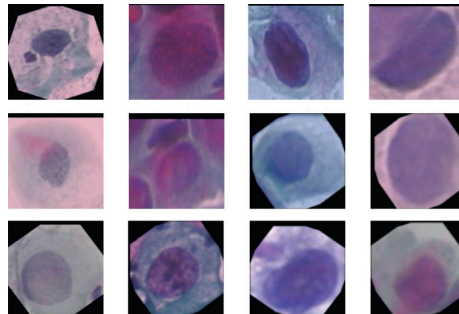


FIGURE 14: Correctly classified malignant cell images.

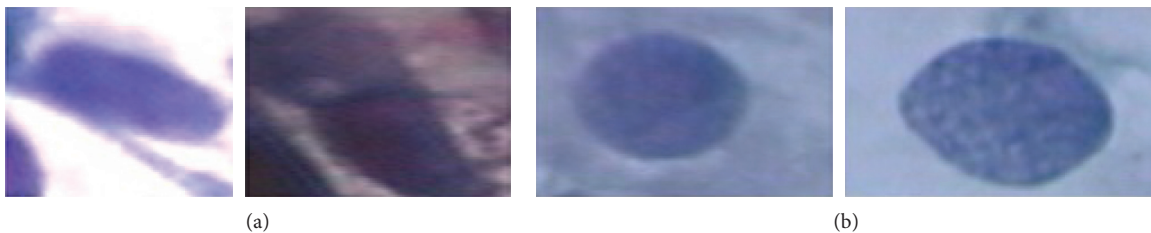


FIGURE 15: Cervical cell images that are misclassified: (a) normal as abnormal; (b) abnormal as normal.

review of the existing methods on cervical cell images such as [13, 39], and [50–56] shows classification performance in terms of accuracy metric only. On the contrary, we have considered four distinct

evaluation metrics including accuracy, sensitivity, specificity, and area under the curve. The experimental results displayed in Tables 5 and 6 show the consistent performance of our proposed models in

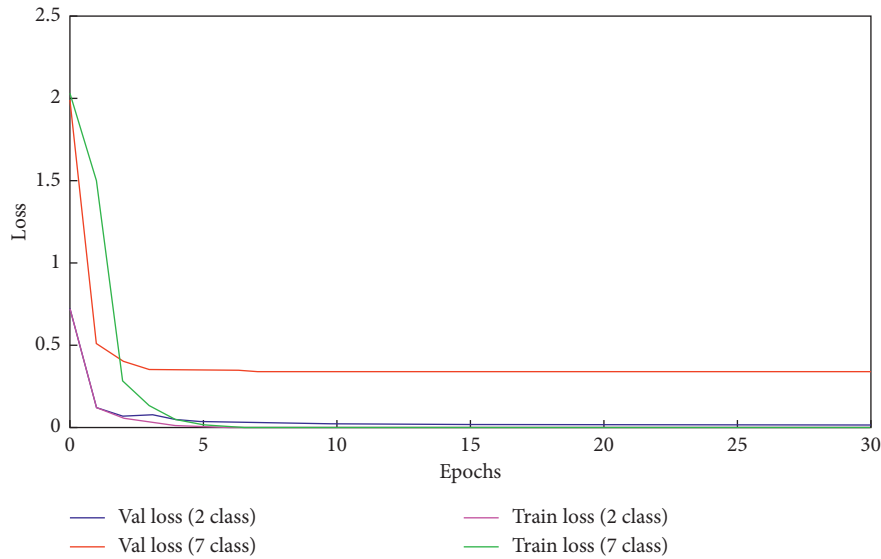


FIGURE 16: Effect of data augmentation on training and validation loss for 2-class and 7-class.

cervical cell images classification across different evaluation metrics. We have proposed two scenarios with different classifiers, i.e., SVM and GEDT. It is noted that the proposed scheme outperforms all previous approaches. This is despite of the fact that training and test images also contained images of overlapping cells. This exceptional performance is mainly because of following reasons: (1) during the training stage, transfer learning is used and the network is finally, fine-tuned on the Herlev dataset; (2) the trained network is used only for extraction of deep features; and (3) the extracted features are then fed to more robust classifiers like SVM and GEDT which are used for final classification. This suggests the effectiveness of our method for underlying task in the presence of a wide variety of cervical cell images ranging from class 1 to class 7.

- (5) The structure of different layers of fine-tuned network is also explored. It is seen that the features learned in the initial *conv* layers are more generalized, and as we move deeper into the network, the extracted features tend to become more abstract. The features learned at fully connected layers are displayed in Figures 9–12.
- (6) Despite higher performance of deep learning-based cervical cell screening system, it has some limitations. Classification time of testing a cropped single cell image is 8 seconds for the system that is very slow in clinical setting as there are large numbers of samples from one PAP smear slide. This limitation can be addressed by neglecting the process of data augmentation step for the test data, and only multiple crop testing can be used for classification problem. This increases the speed of the system as it requires only 0.08 seconds for classification, but the accuracy of the system is compromised by 1.5%. Although classification accuracy of the system on the

Herlev dataset is high, there is room for further improvement.

5. Conclusions and Future Work

This paper proposes an automatic cervical cancer screening system using convolutional neural network. Unlike previous methods, which are based upon cytoplasm/nucleus segmentation and hand-crafted features, our method automatically extracts deep features embedded in the cell image patch for classification. This system requires cells with coarsely centered nucleus as the network input. Transfer learning is used for pretraining, and initial weights or feature maps are transferred from a pretrained network to a new convolutional neural network for fine-tuning on the cervical cell dataset. The features learned by the new fine-tuned network are extracted and given as input to different classifiers, i.e., SR, SVM, and GEDT. The validation results for 2-class and 7-class problems are analyzed. To test a single cell image, different test image patches are generated same as training data, and the multiple crop testing scheme has been carried out on all patches to achieve classifier score. It is further aggregated for the final score. The proposed method yields the highest performance, as compared to previous state-of-the-art approaches in terms of classification accuracy, sensitivity, specificity, F1 score, and area under the curve on the Herlev Pap smear dataset. It is anticipated that a segmentation free, highly accurate cervical cell classification system of this type is a promising approach for the development of auto-assisted cervical cancer screening system.

In future, the effect of system on field of view images containing overlapping cells is to be analyzed. The system should avoid the misclassification of overlapping objects. Specific classifiers relying on deep learning may be used to cater these problems. Moreover, deep learning-based cervical cell classification still needs to be explored for high-precision diagnosis.

Data Availability

The data used to support the findings of this study will be made available on request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] C. Marth, F. Landoni, S. Mahner, M. McCormack, A. Gonzalez-Martin, and N. Colombo, "Cervical cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 28, p. iv72, 2017.
- [2] M. F. Janicek and H. E. Averette, "Cervical cancer: prevention, diagnosis, and therapeutics," *CA: A Cancer Journal for Clinicians*, vol. 51, no. 2, p. 92, 2001.
- [3] R. M. Demay, "Common problems in papanicolaou smear interpretation," *Archives of Pathology and Laboratory Medicine*, vol. 121, no. 3, pp. 229–238, 1997.
- [4] M. Arya, N. Mittal, and G. Singh, "Cervical cancer detection using segmentation on pap smear images," *Proceedings of the International Conference on Informatics and Analytics*, vol. 16, 2016.
- [5] F. Bray, B. Carstensen, H. Møller et al., "Incidence trends of adenocarcinoma of the cervix in 13 European countries," *Cancer Epidemiology Biomarkers & Prevention*, vol. 14, no. 9, pp. 2191–2199, 2005.
- [6] D. Saslow, D. Solomon, H. W. Lawson et al., "American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 3, 2012.
- [7] G. G. Birdsong, "Automated screening of cervical cytology specimens," *Human Pathology*, vol. 27, no. 5, pp. 468–481, 1996.
- [8] H. C. Kitchener, R. Blanks, G. Dunn et al., "Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial," *The Lancet Oncology*, vol. 12, no. 1, pp. 56–64, 2011.
- [9] E. Bengtsson, P. Malm, E. Bengtsson, and P. Malm, "Screening for cervical cancer using automated analysis of PAP-smears," *Computational and Mathematical Methods in Medicine*, vol. 2014, p. 1, 2014.
- [10] L. Zhang, H. Kong, C. Ting Chin et al., "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," *Cytometry Part A*, vol. 85, no. 3, p. 214, 2014.
- [11] E. Bak, K. van Najarian, and J. P. Brockway, "Efficient segmentation framework of cell images in noise environments," in *Proceedings of Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, National Center for Biotechnology Information, Bethesda, MD, USA, July 2004.
- [12] Rahmadwati, G. Naghdy, M. Ros, C. Todd, and E. Norahmawati, "Cervical cancer classification using Gabor filters," in *Proceedings of 2011 1st IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology*, July 2011.
- [13] P. Wang, L. Wang, Y. Li, Q. Song, S. Lv, and X. Hu, "Automatic cell nuclei segmentation and classification of cervical pap smear images," *Biomedical Signal Processing and Control*, vol. 48, pp. 93–103, 2019.
- [14] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in Pap smears," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, p. 539, 2014.
- [15] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images," *BioMedical Engineering Online*, vol. 18, no. 1, pp. 1–22, 2019.
- [16] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, 2011.
- [17] A. Garrido and N. Pérez De La Blanca, "Applying deformable templates for cell image segmentation," *Pattern Recognition*, vol. 33, no. 5, pp. 821–832, 2000.
- [18] R. Saha, M. Bajger, and G. Lee, "Prior guided segmentation and nuclei feature based abnormality detection in cervical cells," in *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019*, pp. 742–746, Institute of Electrical and Electronics Engineers Inc., Athens, Greece, October 2019.
- [19] N. Lavanya Devi and P. Thirumurugan, "Automated detection of cervical cancer," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 1399–1401, 2019.
- [20] M. Jones and L. R. McNally, "A new approach for automated image segmentation of organs at risk in cervical cancer," *Radiology: Imaging Cancer*, vol. 2, no. 2, 2020.
- [21] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proceedings of 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015.
- [22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, 2016.
- [23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [24] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, p. 541, 1989.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances In Neural Information Processing Systems*, vol. 25, no. 2, 2012.
- [26] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [27] R. M. Summers, "Progress in fully automated abdominal CT interpretation," *American Journal of Roentgenology*, vol. 207, no. 1, Article ID 15334406, 2016.
- [28] H.-C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, p. 1285, 2016.
- [29] H. R. Roth, L. Lu, J. Liu et al., "Improving computer-aided detection using_newline convolutional neural networks and

- random view aggregation,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, p. 1170, 2016.
- [30] T. Xu, H. Zhang, C. Xin et al., “Multi-feature based benchmark for cervical dysplasia classification evaluation,” *Pattern Recognition*, vol. 63, pp. 468–475, 2017.
- [31] T. Xu, H. Zhang, X. Huang, S. Zhang, and D. N. Metaxas, “Multimodal deep learning for cervical dysplasia diagnosis,” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9901, pp. 115–123, 2016.
- [32] P. Buysens, A. Elmoataz, and O. Lézoray, “Multiscale convolutional neural networks for vision-based classification of cells,” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7725, pp. 342–352, 2013.
- [33] J. Lemley, S. Bazrafkan, and P. Corcoran, “Smart augmentation learning an optimal data augmentation strategy,” *IEEE Access*, vol. 5, p. 5858, 2017.
- [34] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, “Pap-smear benchmark data for pattern classification,” in *Proceedings of NiSIS 2005: Nature Inspired Smart Information Systems*, January 2005.
- [35] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” 2014, <https://arxiv.org/abs/1411.1792>.
- [36] L. Zhang, L. Le Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, “DeepPap: deep convolutional networks for cervical cell classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, p. 1633, 2017.
- [37] C. Li, D. Xue, X. Zhou et al., “Transfer learning based classification of cervical cancer immunohistochemistry images,” in *Proceedings of the Third International Symposium on Image Computing and Digital Medicine - ISICDM 2019*, pp. 102–106, ACM, Xi’an, China, August 2019.
- [38] K. Bora, M. Chowdhury, L. B. Mahanta, M. K. Kundu, and A. K. Das, “Automated classification of pap smear images to detect cervical dysplasia,” *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 31–47, 2017.
- [39] A. Ghoneim, G. Muhammad, and M. S. Hossain, “Cervical cancer classification using convolutional neural networks and extreme learning machines,” *Future Generation Computer Systems*, vol. 102, pp. 643–649, 2020.
- [40] L. Nanni, A. Lumini, and S. Brahmam, “Local binary patterns variants as texture descriptors for medical image analysis,” *Artificial Intelligence in Medicine*, vol. 49, no. 2, pp. 117–125, 2010.
- [41] Y. Guo, G. Zhao, and M. Pietikäinen, “Discriminative features for texture description,” *Pattern Recognition*, vol. 45, no. 10, pp. 3834–3843, 2012.
- [42] B. Sokouti, S. Haghypour, and A. D. Tabrizi, “A framework for diagnosing cervical cancer disease based on feedforward MLP neural network and ThinPrep histopathological cell image features,” *Neural Computing and Applications*, vol. 24, no. 1, p. 221, 2014.
- [43] Z. Gao, L. Wang, L. Zhou, and J. Zhang, “HEp-2 cell image classification with deep convolutional neural networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, p. 416, 2017.
- [44] Z. Lu, G. Carneiro, A. P. Bradley et al., “Evaluation of three algorithms for the segmentation of overlapping cervical cells,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, p. 441, 2017.
- [45] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, 2009.
- [46] D. Jia, W. Dong., R. Socher, L. Li-Jia, L. Kai, and L. Fei-Fei, “ImageNet: a large-scale hierarchical image database,” in *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.
- [47] O. Russakovsky, J. Deng, H. Su et al., “ImageNet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, p. 211, 2015.
- [48] A. Moayedikia, K. L. Ong, Y. L. Boo, W. G. Yeoh, and R. Jensen, “Feature selection for high dimensional imbalanced class data using harmony search,” *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 38–49, 2017.
- [49] D. Powers, “Evaluation: from precision, recall and F-measure to roc, informedness, markedness & correlation,” *Journal of Machine Learning Technologies*, vol. 35, no. 1, pp. 41–64, 2011.
- [50] J. Zhang and Y. Liu, “Cervical cancer detection using SVM based feature screening,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 3217, pp. 873–880, 2004.
- [51] P. K. Malli and S. Nandyal, “Machine learning technique for detection of cervical cancer using k-NN and artificial neural network,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 6, 2017.
- [52] N. Ampazis, G. Dounias, and J. Jantzen, “Pap-smear classification using efficient second order neural network training algorithms,” in *Proceeding of the Hellenic Conference on Artificial Intelligence*, pp. 230–245, Samos, Greece, May 2004.
- [53] S. Sharma, “Cervical cancer stage prediction using decision tree approach of machine learning,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, pp. 345–348, 2016.
- [54] R. Gupta, A. Sarwar, and V. Sharma, “Screening of cervical cancer by artificial intelligence based analysis of digitized papanicolaou-smear images,” *International Journal of Contemporary Medical Research*, vol. 4, 2017.
- [55] G. Dounias, B. Bjerregaard, J. Jantzen et al., “Automated identification of cancerous smears using various competitive intelligent techniques,” *Oncology Reports*, vol. 15, pp. 1001–1006, 2006.
- [56] Y. Marinakis, G. Dounias, and J. Jantzen, “Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification,” *Computers in Biology and Medicine*, vol. 39, no. 1, p. 69, 2009.