

Pap Smear Image Classification Using Convolutional Neural Network

Kangkana Bora
Central Computational and
Numerical Studies
Institute of Advanced Study in
Science and Technology
Guwahati, Assam, India
kangkana.bora89@gmail.com

Manish Chowdhury
KTH, School of Technology
and Health,
Stockholm, Sweden
st.manishc@gmail.com

Lipi B. Mahanta
Central Computational and
Numerical Studies
Institute of Advanced Study in
Science and Technology
Guwahati, Assam, India
lipimahanta@yahoo.co.in

Malay K. Kundu
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
malay@isical.ac.in

Anup K. Das
Department of Pathology
Ayursundra Healthcare Pvt.
Ltd.
Guwahati, Assam, India
anup.das@ayursundra.in

ABSTRACT

This article presents the result of a comprehensive study on deep learning based Computer Aided Diagnostic techniques for classification of cervical dysplasia using Pap smear images. All the experiments are performed on a real indigenously image database containing 1611 images, generated at two diagnostic centres. Focus is given on constructing an effective feature vector which can perform multiple level of representation of the features hidden in a Pap smear image. For this purpose Deep Convolutional Neural Network is used, followed by feature selection using an unsupervised technique with Maximal Information Compression Index as similarity measure. Finally performance of two classifiers namely Least Square Support Vector Machine (LSSVM) and Softmax Regression are monitored and classifier selection is performed based on five measures along with five fold cross validation technique. Output classes reflects the established Bethesda system of classification for identifying pre-cancerous and cancerous lesion of cervix. The proposed system is also compared with two existing conventional systems and also tested on a publicly available database. Experimental results and comparison shows that proposed system performs efficiently in Pap smear classification.

Keywords

Pap smear image; Deep learning; LSSVM; Softmax Regression

1. INTRODUCTION

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

ICVGIP, December 18-22, 2016, Guwahati, India

© 2016 ACM. ISBN 978-1-4503-4753-2/16/12...\$15.00

DOI: <http://dx.doi.org/10.1145/3009977.3010068>

Computer Aided Diagnosis (CADx) is a technique designed to decrease observational oversights and thus the false negative rates of physicians interpreting medical images [3]. Recent advances in application of CADx systems in medical field increases the performance of decision delivery by removing inter-observer variations and providing quantitative support in clinical decision making [6]. Feature extraction, feature selection and classification are three major phases of designing any such CADx system. Efficiency of any automated system is fully determined by the above mentioned techniques. In recent years deep learning techniques has made a good impact on medical image processing and developing of CADx systems [1][6]. Specially Convolutional Neural Network (CNN) has become popular for many computer vision works. But limitations exists when it comes to the application on cytological image processing, because to train a deep network huge amount of data is required. But as mentioned by Genctav *et al.*, [8] collection of data is a big challenge in medical imaging field.

Attempt has been made to study the features of Pap smear images through segmentation techniques like FCM [4], Genetic Algorithm [17], Joint Optimization technique [19] and watershed transform [22, 21] etc. Authors have first tried to segment the region of interest (nucleus or cytoplasm) and tried to extract the morphometric or textural features from the region of interest. After extraction of features different classifiers were used to classify the Pap smear images [2, 4, 12]. Limiting factor of such quantitative image analysis technique is the absence of a robust and ideal segmentation algorithm to distinguish object of interest [5]. Also one segmentation technique may not be consistent in segmenting images collected from different sources. To overcome these drawbacks we are using Deep features extracted using CNN. Deep learning methods are a set of algorithms in machine learning, which try to automatically learn multiple levels of representation and abstraction of images that help make sense of data [24].

In this work attempt has been made to classify Pap smear single cell images to detect cervical dysplasia. All the experiments were performed on a generated database containing

1611 images of different categories namely Negative for Intraepithelial Lesion or Malignancy (NILM), Low-grade Squamous Intraepithelial Lesion (LSIL) and High-grade Squamous Intraepithelial Lesion or HSIL (including Squamous Cell Carcinoma). These categories reflect the established Bethesda System of classification for identifying cancerous and pre-cancerous lesion of cervix. System is also tested on an online Herlev database containing 917 images which is publicly available.¹ Feature extraction is performed using Deep CNN. For feature selection unsupervised method based on similarity measure is considered so that we can remove the redundant features to improve accuracy of the system. Finally performance of two classifiers namely Least Square Support Vector Machine (LSSVM) and softmax regression are monitored, followed by classifier selection based on 5 measures namely Accuracy, Precision, Recall, Specificity and F-Score. Five fold cross validation is used for assessment purpose. Output classes will reflect the degree of dysplasia present in an image. The system is also compared with two existing conventional methods for Pap smear classification. Extensive experiments proved that the proposed system outperform the existing techniques.

Main contribution of the paper can be summarized as follows - i) *Database generation*: A new Pap smear image database is generated to perform all the experiments where images are collected from two diagnostic centers. Genctav *et al* [8] mentioned the difficulty in collecting real indigenous samples for experimental purpose. To avoid ethical issues involved in this type of studies researchers perform their study on publicly available data. But the system developed on those data may not show consistent result in region based real samples. ii) *In depth analysis of deep learning features using CNN*: Although literature is rich in studying the application of deep learning in various domains but hardly any study reported for application of deep learning in Pap smear classification. This motivates the current work to use deep features in cervical dysplasia detection.

2. METHODS

The block diagram of the proposed work is shown in Fig 1. It involves 4 phases. Phase I is for database generation, Phase II is for feature extraction using deep CNN, Phase III is for feature selection using unsupervised technique and final phase is for classification where output classes reflects the established pathological classification system.

2.1 Database Generation

1611 real Pap smear images were collected from Ayursundra Healthcare Pvt. Ltd and Dr. B. Borooah Cancer Institute, Guwahati, Assam. Staining and preparation of slides were done at the respective centers by certified pathologist. Then the images were captured by us using Leica ICC50 HD microscope under 400X resolution with 24bit color depth. The images were labeled based on the report collected from the diagnostic centers. Finally an image database is prepared with the help of two doctors from the respective diagnostic centers. Characteristic of NILM, LSIL and HSIL is well explained in the book of Gray [9]. A sample of images of the database is displayed in Fig 2. To check the consistency of the system same set of experiments were performed

¹The download link of Herlev University Pap smear database is <http://labs.fme.aegean.gr/decision/downloads>

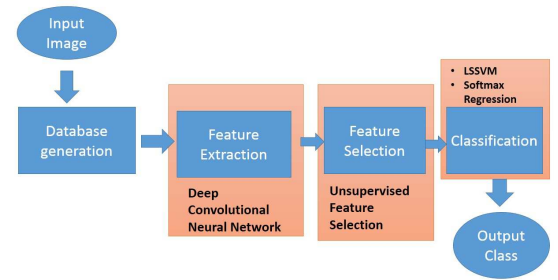


Figure 1: Block diagram of the proposed work

on Herlev University Pap smear database which is publicly available. Details of both the database is available in Table 1. Since in Herlev database images were divided into 7 categories we have grouped them into 3 categories according to Bethesda standard where Normal Intermediate, Normal Superficial and Normal Columnar were included in NILM category. Mild dysplasia is included in LSIL and other 2 categories were included in HSIL class. So final classes are as follows- Class 1: NILM, Class 2:LSIL and Class 3: HSIL including Squamous Cell Carcinoma.

Table 1: Details of images collected

Own database: Cell level	Details
Resolution	400X
NILM	1001 (Class 1 image)
LSIL	400 (Class 2 image)
HSIL including Squamous cell carcinoma	210(Class 3 images)
Total image	1611
Herlev University database	
Normal Intermediate	70
Normal Superficial	74
Normal Columnar	98
Mild dysplasia	182
Moderate dysplasia	146
Severe dysplasia	197
Carcinoma in situ	150
Total image	917

2.2 Feature Extraction using CNN

The performance of the image retrieval system is inherently depends on the effectiveness of the feature vector representing content of the image. Recent result signify that the generic descriptors extracted from CNNs are exceptionally effective in object recognition and currently the hot discussion in computer vision forum [14] [18]. CNNs comprise a feed-forward family of deep networks, where in-between layers receive as input the features generated by the previous layer, and pass their outputs to the subsequently layer. The strength of this network is in learning hierarchical layers of concept representation, corresponding to different levels of abstraction. For image data, the lower levels of abstraction might describe the different orientated edges in the image; middle levels might describe parts of an object while high

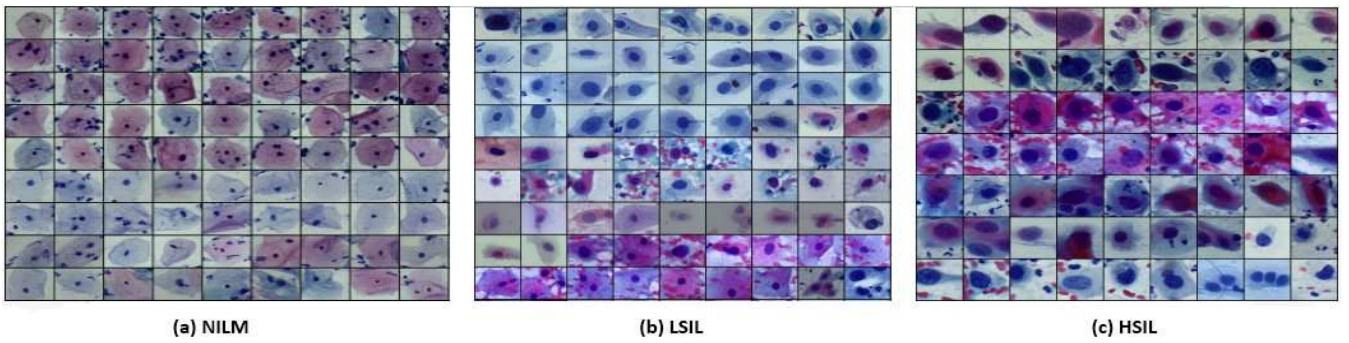


Figure 2: Sample images of different categories

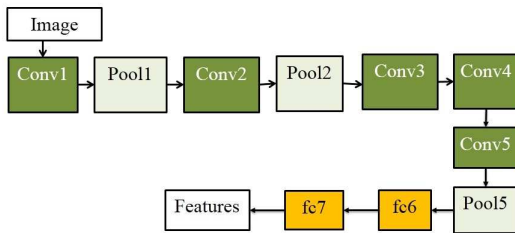


Figure 3: Block Diagram of AlexNet CNN Model

layers refer to larger object parts and even the object itself. Deep learning methods are most efficient when applied on large training sets [16]. In medical imaging research, such huge data-set are difficult to get [8].

Few recent studies can be found in the medical field that uses deep architecture methods [1, 23]. In [1], Imagenet non-medical trained model has been used for classification of medical images. Using same intuition, we have used Berkeley Caffe reference model imagenet-caffe-alex [13] for pap smear image features extraction mechanism. AlexNet has five convolution layers, three pooling layers, and two fully-connected layers with approximately 60 million free parameters [25]. For our work, we have chosen this network because AlexNet is originally designed for the fixed image dimension of $256 \times 256 \times 3$ pixels [15]. This model is also giving good performance when the object of interest in small and obscure [25]. In our case, identification of nucleus boundary is pose more challenges for learning a successful classification model. Fig 3 shows the block diagram of AlexNet model.

For this paper we have trained and tested the model from scratch and no pre-trained model is used for this purpose. System is individually trained and tested for generated single cell and Herlev database. During the process, the image of the medical database are re-sized to $M \times M$. These re-sized images are used to construct the image representative feature vectors. CNNs comprise a feed forward family of deep networks, where in- between layers receive as input the features generated by the previous layer, and pass their outputs to the subsequent layers. The strength of this network is in learning hierarchical layers of concept representation, corresponding to different levels of abstraction. We choose the output of the 'fc7' layer as the feature vector. The rea-

son we choose this layer against other layer is because the 'fc7' is the last layer before the final output layer (the layer which output the class scores) and should contain more specific details of the images. Finally each image is represented with 4,096 dimensional feature vectors..

2.3 Feature Selection

Selection of a smaller subset of features, retaining the optimal characteristics of the data, not only reduces the computational time and complexity but also lead to generation of a compact efficient model. In this paper we are trying to use an unsupervised feature selection technique proposed by Mitra *et al* [20]. It is based on measuring similarity using Maximal Information Compression Index (MICI) among feature vectors. It is fast as it does not require any search and feature subset is not transformed unlike Principal Component Analysis (PCA). In involves partitioning a feature set into smaller subsets or clusters where feature within a cluster are highly similar but inter cluster similarity is very low. A single feature from each such cluster is then selected to constitute the resulting reduced subset. Detailed algorithm is available in [20].

2.4 Classification

In this paper two classifiers performance is monitored namely LSSVM and Softmax Regression and classifier selection is performed based on 5 measures stated in Table 2.

2.4.1 LSSVM

The most critical drawback of Support Vector Machine (SVM) is its high computational complexity for high dimensional data sets. To reduce the computational demand, the least square version of SVM (LSSVM) is adopted as classifier in this paper. LSSVM avoids solving quadratic programming problem and simplifies the training procedure [7].

LS-SVM was originally developed for binary classification problems. But, a number of methods have been proposed by various researchers for extension of binary classification to multi-classification problem. However, one need an appropriate method for solving this multi-class problem. It's essentially separate M mutually exclusive classes by solving many two-class problems and combining their predictions in various ways. One such technique which is commonly used is Pairwise Coupling (PWC) or "one-vs.-one" is to construct binary SVMs between all possible pairs of classes. PWC uses $M * (M - 1) / 2$ binary classifiers for M number of classes, each of which provides a partial decision for clas-

sifying a data point. During the testing of a feature, each of the $M * (M - 1)/2$ classifiers votes for one class. The winning class is the one with the largest number of accumulated votes. Hsu *et al.*, shows that the PWC method is more suitable for practical use than the other methods discussed in [11]. Hence, we use the one against one multi-class image classification method based on the LSSVM tool by combining all pairwise comparison of binary classifiers. The details of the LSSVM is discussed in [27].

2.4.2 Softmax Regression

Softmax regression is a generalization of logistic regression where we deal with multiple classes. The label $y^{(i)} = \{1, 2, \dots, C\}$ where C represent the number of classes [10]. Given a test input x , it finds the hypothesis to estimate the probability that $P(y = k|x)$ for each value of $c = 1, 2, \dots, C$. Thus the hypothesis will output a C -dimensional vector giving C estimated probabilities. The softmax function ς takes an input C - dimensional vector z and outputs a C - dimensional vector y of the real values between 0 and 1. This function is a normalized exponential and is defined as:

$$y_c = \varsigma(z)_c = \frac{e^{z_c}}{\sum_{d=1}^C e^{z_d}} \text{ for } c = 1, 2, \dots, C \quad (1)$$

The denominator $\sum_{d=1}^C e^{z_d}$ acts as a regularizer to make sure that $\sum_{c=1}^C y_c = 1$. We can write the probabilities that the class $t = c$ for $c = 1 \dots C$ given input z as:

$$P = \varsigma = \frac{1}{\sum_{d=1}^C e^{z_d}} E \quad (2)$$

where,

$$P = \begin{bmatrix} P(t = 1|z) \\ \vdots \\ P(t = C|z) \end{bmatrix}$$

$$\varsigma = \begin{bmatrix} \varsigma(z)_1 \\ \vdots \\ \varsigma(z)_C \end{bmatrix}$$

$$E = \begin{bmatrix} e^{z_1} \\ \vdots \\ e^{z_C} \end{bmatrix}$$

$P(t = c|z)$ is thus the probability that the class is c given the input z .

3. RESULT AND DISCUSSION

3.1 Experimental set-up

Implementations were carried out in MATLAB R2016a using Intel CORE i5 processor of 2.20GHz and 4GB RAM. First an image repository was created where all the collected images were stored in 3 different directories which represents 3 different classes. Images were arranged according to the reports of the patients collected and with proper consultation with the pathologist. For assessment of the technique 5 measures are taken into consideration. These measures are listed in Table 2.

The accuracy of classification was evaluated by computing the number of correctly recognized class examples (true positives), the number of correctly recognized samples that do not belong to the class (true negatives) and examples that

Table 2: Measures for efficiency evaluation of classification tp_i =true positive of class C_i , tn_i =true negative of class C_i , fp_i =false positive of class C_i , fn_i =false negative of class C_i , l =number of classes

Assessments	Fomulae
Average accuracy	$\frac{\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l}$
Precision	$\frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i}}{l}$
Recall	$\frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i}}{l}$
Specificity	$\frac{\sum_{i=1}^l \frac{tn_i}{tn_i + fp_i}}{l}$
F score	$2 * \frac{Precision * Recall}{Precision + Recall}$

either were incorrectly assigned to the class (false positives) or that were not recognized as class examples (false negatives) [26]. The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. Higher the accuracy, higher the rate of truly classified classes. The value of recall reflects the rate of miss-classification i.e., higher the value of recall, the lower will be the rate of misclassification. On the other hand, precision reflects the rate of false positives; lower is the rate of false positives, higher is the precision. F-measure is a harmonic mean of recall and precision. The value of F-measure will be high only when both the recall and precisions are high, i.e. only when the miss-classification rates as well as rate of false positives both are low.

3.2 Evaluation of feature selection technique

Since the feature vector extracted using CNN is of high dimensional (of size 4096) that is why feature selection is performed to check the presence of redundant features and remove them to increase the efficiency of the system. The feature selection technique used in this approach involves two steps. Firstly, it partition the original feature set into some clusters using k NN principle using MICI as similarity measure [20]. Secondly, selecting a representative feature from each such cluster. In this approach size of reduced feature is totally determined by the parameter k . To perform the experiments and choosing of the effective reduced feature size R from the original feature size D a small experiment is performed. In doing so k values have chosen in such a way that $R = 100, 200, 300, \dots, 4000$. For each reduced feature subset the accuracy value using LSSVM classifier is monitored. From 1611 images collected 70% of the images were considered as training set and rest 30% were considered as test set. The R value giving maximum accuracy has chosen as the final R and used in further processing. Fig 4 shows accuracy vs reduced feature set having different R value where $R = 100, 200, \dots, 4000$. From Fig 4 it can be observed that for $R = 100$ the accuracy is maximum, hence the final reduced feature set size is considered as 100. It was also observed from Fig 4 that higher dimensional features of size 4096 is giving accuracy in the rage of 84-87% for generated database. It indicate the presence of redundant features in the feature vector. So after application of feature selection accuracy improved to 90-95% which also lowers the dimension to 100, it indicate the importance of using this technique. It can also reduce the computational

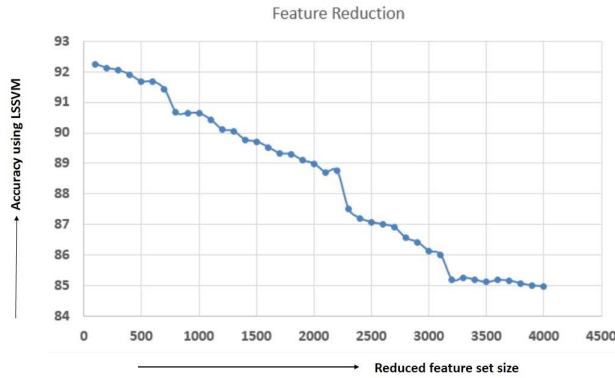


Figure 4: Selection of reduced feature set dimension

time of the system.

3.3 Evaluation of the classification techniques

Evaluation is performed in two phase. One before feature selection and other one is done after feature selection so that we can monitor the performance as accurately as possible. For evaluation purpose 5 measures stated in Table 2 were consideration namely Accuracy, Precision, Recall, Specificity and F-Score. Experiments were performed on generated database as well as on Herlev database individually. Classification is of two level - 2 class and 3 class. In 2 level of classification the output classes were Normal and Abnormal classes, where NILM is included in Normal class and LSIL with HSIL is included in Abnormal class, abnormal class also includes the Squamous carcinoma cases. On the other hand, in 3 level of classification the output classes are NILM, LSIL and HSIL (including squamous cell carcinoma). For 2 level classification Softmax regression is not applicable as it can be applied in multiclass problem only, therefore logistic regression is used in place of softmax regression for 2 class problems. Fig 5 displays different measures of the proposed work before feature selection and Fig 6 displays the result of the system after feature selection. It can be observed that the result of system after feature selection is very much satisfactory. 5 fold cross validation is performed to achieve good assessment of the system. K fold cross validation is one of the best techniques for training and testing a given module to increase the performance of the system on the data already available. In this case dataset is divided into k subsets and the method repeated k times. Each time one of the k subsets is used as test set and other $k - 1$ subsets are considered as training set. The advantage of this technique is that the technique does not affected by the matter how the dataset are divided. It can represent the given dataset nicely. Using this technique the developer is independent of choosing the size of train and test set based on K- folds used. Here to get the advantage of this technique also keeping in mind that it should not increase the computational time to an extent because of the repetition performed, we choose $k = 5$ for computation purpose.

It can be observed from the results of Fig 6 Softmax Regression perform better then LSSVM in case of 3 level classification. To delve deeper into the study next set of experiments were performed to check whether the adopted model

using softmax regression fit the data or not. For this purpose statistical approaches were followed to compute the goodness of fit using Pearson chi-square analysis. Along with it model fitting information were observed carefully to check its significance. The observed value of dF and p is displayed in Table 3. p value 'indicate 1% level of significance' when $p < 0.01$ and when $p < 0.05$ it 'indicate 5% level of significance'. On the other hand, higher the value of dF more efficient is the model. Values obtained from Table 3 proved that the model very much fit the data and softmax regression can be efficiently use in 3 level classification.

Table 3: Result of Statistical Analysis

Model fitting information		
Database used	dF	p
Generated	202	.000
Herlev	200	.000
Goodness-of-fit		
Generated	3018	.000
Herlev	1630	.050

3.4 Comparison with two existing technique specifically applied to Pap smear image classification

Next set of experiments were performed to further justify our work by comparing the proposed work with two existing conventional techniques of Pap smear classification [4][12]. The researchers of both paper performed their experiment on Herlev database. But to check consistency of the system developed and for deliverable output to society one must test their system with real time data. In [4], authors have tried to extract 9 features including morphometric and textural features. In [12], 20 features were extracted from pap smear images and all of them describe morphology (such as size and shape) of cervical cell. Then classified using some known techniques like SVM, KNN etc. They have basically followed segmentation based approaches to extract the features. But the main disadvantage of segmentation based approaches is that they are not perfectly ideal. No segmentation technique can give 100% accuracy under any circumstances. It can be observed from the results of Fig 7,8 that proposed system outperform both the techniques using LSSVM or Softmax as classifier and deep CNN as feature vector design principle. The proposed system directly extract Deep features using CNN without segmenting the region of interest. So one of the advantage of this technique is that it does not incorporate the disadvantage of segmentation based approaches. Also CNN can perform multiple level of representation of an image and extract best possible features hidden in an image which is not possible using conventional techniques.

One drawback of Deep learning based technique is that its training time is high as compared to conventional technique and this mainly arises because of larger feature set dimension. This high dimensional feature may contain some redundant information as in this case, where classification result with 4096 features is in the range of 84-87% as shown in Fig 5. But after application of feature selection technique accuracy improved to the range 90-95% (Fig 6). In future if we can increase the training sample size then deep learning based technique will outperform all the existing techniques.

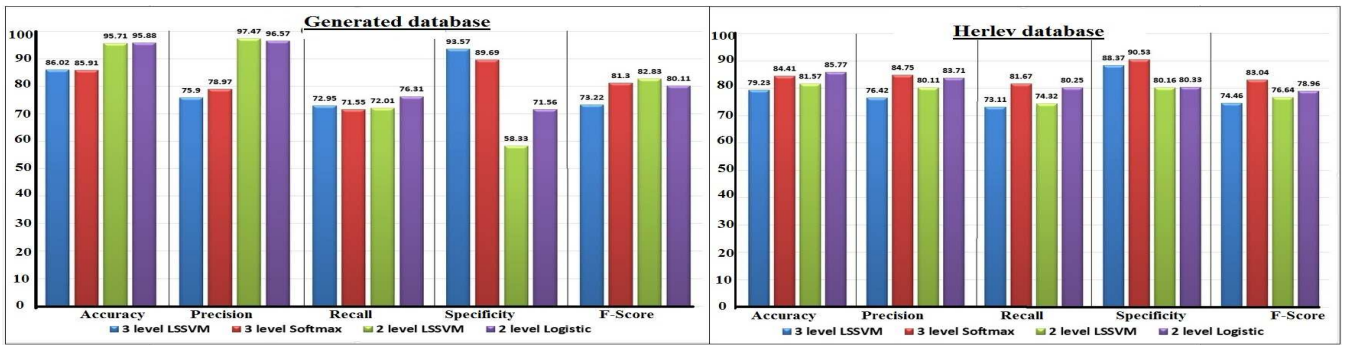


Figure 5: Result of classification before feature selection using both databases

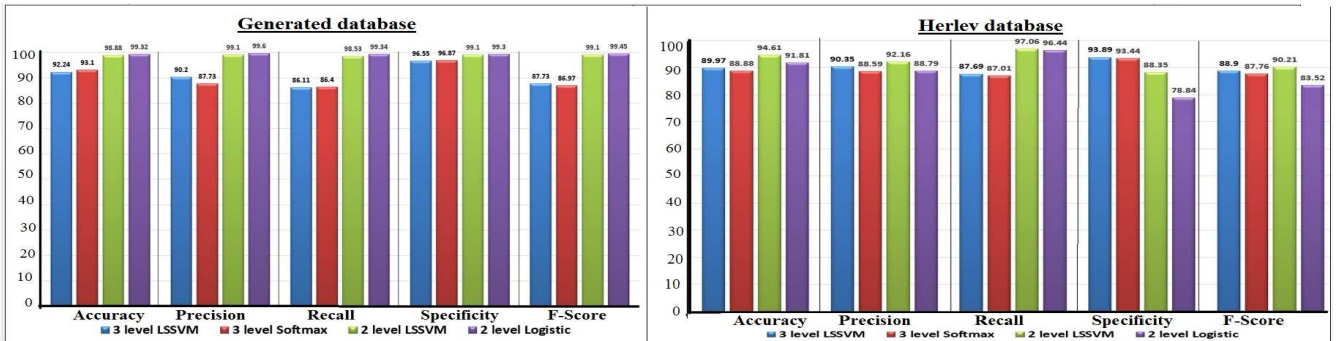


Figure 6: Result of classification after feature selection using both databases

4. CONCLUSIONS

With the increase in the number of patients, pathologist face a substantial increase in the workload. CADx systems can play an important role in reducing false negative cases so that they can concentrate more on suspicious cases. In this paper we have introduced ‘deep learning’ as a tool for efficient and accurate identification of dysplasia from Pap smear images. This type of automated system can help in early detection of cervical cancer. Main importance is given on constructing an effective feature vector representing the image’s content. Experimental evaluation has proved that the proposed system is efficient and improves classification performance. Objective of such system is to increase efficiency by reducing inter-observer variations. Proper training and testing of the system on databases collected from different sources will surely increase the applicability of the system in medical imaging field. Further studies will include more combination of classifiers and feature selection techniques to improve the accuracy of the system.

5. REFERENCES

- [1] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan. Chest pathology detection using deep learning with non-medical training. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 294–297, 2015.
- [2] L. H. Camargo, G. Diaz, and E. Romero. Pap smear cell image classification using global mpeg7 descriptor. *Diagnostic Pathology*, 8((Suppl 1):S38):1–4, 2013.
- [3] R. A. Castellino. Computer aided detection (cad):an overview. *Cancer Imaging*, 5(1):17–19, 2005.
- [4] T. Chankong, N. T. Umpon, and S. Auephanwiriyankul. Automatic cervical cell segmentation and classification in pap smears. *Computer Methods and Programs in Biomedicine*, 113(2):539–556, 2014.
- [5] J. M. Chen, A. P. Qu, L. W. Wang, J. P. Yuan, F. Yang, X. Q. Ming, N. Maskey, G. F. Yang, J. Liu, and Y. Li. New breast cancer prognostic factor identified by computer aided image analysis of he stained histopathology images. *Scientific Reports*, 5:1–13, 2015.
- [6] J. Z. Cheng, D. Ni, J. Qin, C. M. Tiu, Y. C. Chang, C. S. Huang, S. Dinggang, and C. M. Chen. Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific Reports*, 6:1–13, 2016.
- [7] M. Chowdhury and M. K. Kundu. Comparative assessment of efficiency for content based image retrieval systems using different wavelet features and pre-classifier. *Multimedia Tools and Application*, 74(24):11595–11630, 2015.
- [8] A. Genctav, S. Aksoy, and S. Onder. Unsupervised segmentation and classification of cervical cell images. *Pattern Recognition*, 45(12):4151–4168, 2012.
- [9] W. Gray. *Diagnostic Cytopathology*. Churchill Livingstone, 2010.
- [10] D. Gujarati. *Econometrics by Example*. Palgrave, 2012.
- [11] C. W. Hsu and C. J. Lin. A comparison of methods for multi-class support vector machines. *Neural Network, IEEE transaction on*, 13(2):427–437, 2002.

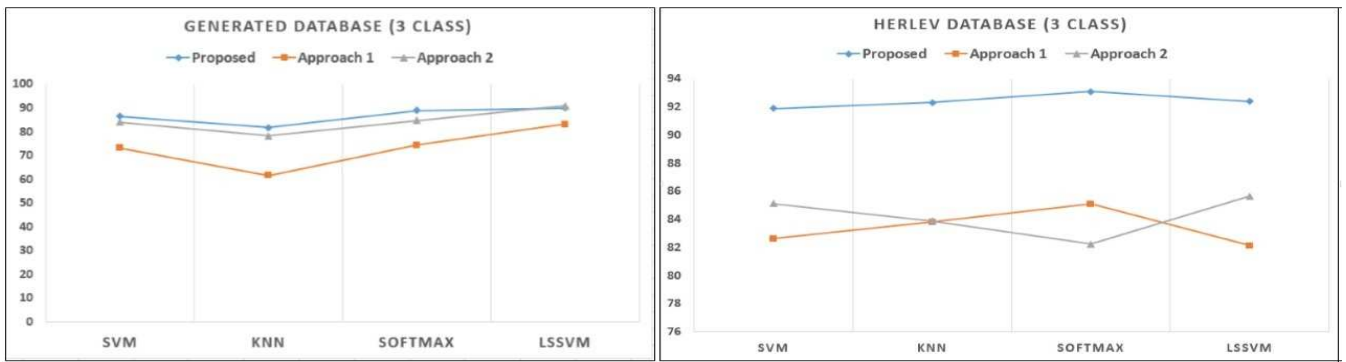


Figure 7: Comparison of the proposed technique with Approach 1 [4] and Approach 2 [12] using 3 level of classification

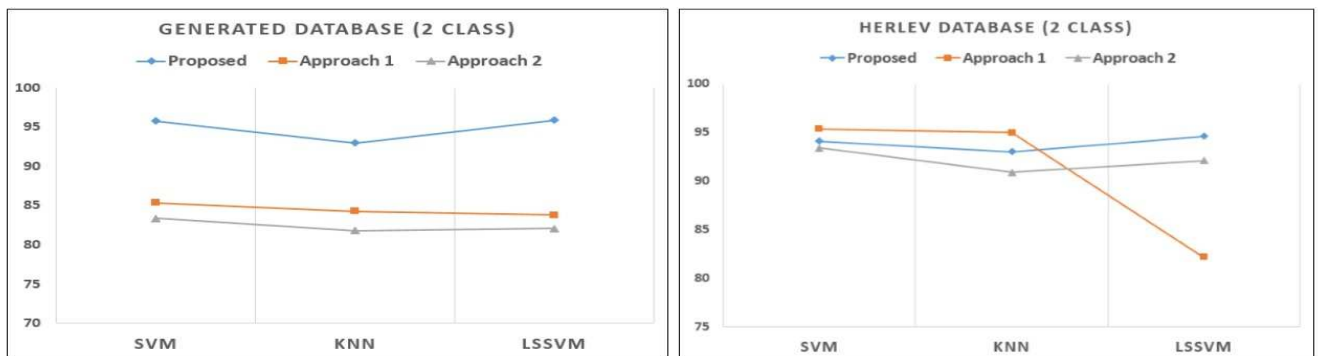


Figure 8: Comparison of the proposed technique with Approach 1 [4] and Approach 2 [12] using 2 level of classification

- [12] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard. Pap-smear benchmark data for pattern classification. In *Proc. NiSIS 2005*, pages 1–9. NiSIS, 2005.
- [13] Y. Jia, E. Shelhamer, J. Donahue, and S. Karayev. Caffe: Convolutional architecture for fast feature embedding. In *In Proc. of MM*, page 675–678, 2016.
- [14] P. Kotschieder, M. Fiterau, A. Criminisi, and S. R. Bulu. Deep neural decision forest. In *Proceeding of ICCV*, pages 1467–1475, 2015.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 1097–1105, YEAR =2012.
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural network. In *Proceeding of NIPS*, pages 1106–1114, 2015.
- [17] N. Lassouaoui, L. Hamami, and N. Nouali. Morphological description of cervical cell images for the pathological recognition. *International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering*, 1(5):307–310, 2007.
- [18] Y. LeCun, Y. Bengioand, and G. E. Hinton. Deep learning. *Nature*, 521:436–444, 2015.
- [19] Z. Lu, G. Carneiro, , and A. P. Bardley. An improved joint optimization of multiple level set function for the segmentation of overlapping cervical cells. *Image Processing, IEEE Transaction on*, 24(4):1261–1272, 2015.
- [20] P. Mitra, C. A. Murthy, and S. K. Pal. Unsupervised feature selection using feature similarity. *Pattern analysis and Machine Inteligence, IEEE Transaction on*, 24(3):301–312, 2002.
- [21] M. E. Plissiti and C. Nikou. Combining shape, texture and intensity features for cell nuclei extraction in pap smear images. *Pattern Recognition Letters*, 32(6):838–853, 2011.
- [22] M. E. Plissiti, C. Nikou, and A. Charchanti. Watershed-based segmentation of cell nuclei boundaries in pap smear images. In *Information Technology and Applications in Biomedicine (ITAB), 10th IEEE International Conference on*, pages 1–4. IEEE, 2010.
- [23] R. Samala, H. P. Chan, L. Hadjiiski, and M. Helvie. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In *SPIE Medical Imaging, Volume: 9785*, 2016.
- [24] P. Seebock. Deep learning in medical image analysis. Master’s thesis, Vienna University of Technology, Faculty of Informatics, 2015.
- [25] H. Shin, H. Roth, M. Chen, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. Summers. Deep

convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *Medical Imaging, IEEE Transaction on*, 2016.

- [26] M. Sokolova and G. Lapalme. A systematic analysis of performance measure for classification task. *Information Processing and Management*, 45(4):427–437, 2009.
- [27] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.