Meticulous review on Cutting-Edge Cervical Cancer cell Detection, Segmentation and Stratification of PapSmear Images using Image Processing and Machine learning Approach

Barkha Bhavsar [1] and Bela Shrimali [2]

[1] *LDRP Institute of Technology and Research, Kadi Sarva Vishwavidyalaya,SVKM,India*
[2] *Assistant Professor, Institute of Technology, Nirma University, India*

Abstract—Cervical cancer falls under the top most cancers found in women of developing countries since last many years. Classification of cervical cancer through a traditional microscopic approach is a monotonous and prolonged task. Most of the time hospital doctors cannot identify the cancer cells as sometimes the nucleus is difficult to see with naked eyes. Due to the different perspectives of doctors, cancer stages are classified falsely which leads to low recovery and late medication. The use of Image Processing and Machine Learning technologies can take off misclassification and inaccurate prediction. Although many deep learning techniques are available for cervical cancer cell detection and classification, performance of such techniques for prediction and classification with the real and sample dataset is the main challenge. In this paper, we did a thorough state-of-the-art review with the available current literature. The objective of this paper is to bring forth in-depth knowledge to novice researchers with the thorough understanding of the architecture of the computer assisted classification process. The current literature is studied, analyzed, and discussed with their approaches, results, and methodologies.

Index Terms—Image Processing, Machine Learning, Image Classification, Pattern Analysis, Feature evaluation, Feature selection.

1 Introduction

Cancer remains the leading causes of death among women across the globe. There is a huge number of cervical cancer victims in developing countries like India and China [1]. According to the World Health Organization (WHO), breast cancer and cervical cancer are common in women compare to other cancers. Based on the Statistics derived from the WHO Global Health Estimates-2019 [2], [3], [4], out of 656 300 000 total population of females in India, 4 191 000 females died due to cervical cancer [5]. In spite of the fact that cervical cancer is one such category of cancer which can be prevented easily through vaccination and also has a very long premalignant period [6], [7] which can be utilized for detection and treatment of the disease, lack of awareness is the key reason for this mortality rate in developing countries [7], [8], [9], [10], [11]. Cervical cancer is developed in the cervix area of vagina the entrance of the uterus in women [1]. HPV vaccination and prevention actions such as regular screening and detection of malignant cells in the cervix help to reduce mortality ratio occurring due to cervical cancer [6], [7], [12], [13].

Primordial diagnosis and screening are effective key factors to prevent cervical cancer effectively [4], [10], [13], [14]. There are various detection and/or screening techniques used to detect precancerous changes that develop in the cervix area over a span of 10 to 20 years [4], [15]. For a single patient the cyto technicians have to examine plenty of the slides of smear under microscope to conclude the accurate result [16]. The conventional approach of screening through visual examination is comparatively time consuming and has a likelyhood of errors. Moreover, lack of trained cytologists and laboratory instruments may lead the whole screening process to cumbersome [8], [10], [16]. This diagnostic screening can be performed either at cellular-level or tissue-level [17]. Pap smear, liquid based cytology (LBC), HPV DNA testing, electromagnetic spectroscopies belong to cellular-level screening approach whereas cervicography, colposcopy, hyperspectral diagnostic imaging (HSDI), and visual inspection after applying acetic acid (VIA) or Lugol's iodine (VILI) belong to tissue-based screening techniques [7], [15], [17], [18], [19], [20], [21], [22]. A Pap test is the most commonly used, easy, quick, painless screening method for cancer and pre-cancer of the uterine cervix [4]. Along with above mentioned issues, these techniques also have a lack of accuracy and informations [10], [16]. Hence , computer assisted screening methods are being introduced to speed up and accurate the whole diagnostic process.

Technological advancement in medical image processing (IP) has proved to be an add-on blessing to the healthcare sector [20]. Medical images captured through various technologies such as X-rays, CT scans, MRI, sonography, Ultrasound, etcetera provides an insight into the patients' body without cutting it open [20], [23], [24], [25]. Detection and identification of the actual problem,

E-mail:barkha.bhavsar@gmail.com
E-mail:bela.shrimali@gmail.com

identifying the root cause of the disease, classification between benign and malignant cells within the human body and suggesting immediate or primary care treatment have become effortless nowadays. Medical image analysis techniques like enhancement, denoising, segmentation, feature extraction, and morphological image processing provides detailed insights into the area of interest due to which doctors and medical team can proceed with the treatment in no time [1], [3], [4], [20]. IP has a wide spectrum of medical applications viz. brain tumor detection, multiple stone detection, identifying fractures, identifying congenital heart defects, breast cancer detection, diagnosing heart valve diseases, tuberculosis detection, identifying birth defects can be measured into different domains where images are used [23], [24], [25]. Particularly, for the successful detection of cervical cancer, Pap smear image analysis has gained a much popularity. In underdeveloped and devloping countries, due to lack of expertise, resources, awareness and remote location,Machine learning has become very popular for medical diagnosis these days [7], [26]. Machine generated results for identifying cervical cancer in the patient prove to be boon to remote locations and/or low income regions. However, it fails in accuracy sometimes due to unfit image having low resolution, noise, overlapping of cervix cells, and texture variation [10].

Machine learning [5], [9], [26], [27] a branch of Artificial Intelligence, which is associated with identifying problems and solutions from the data sample available [16]. These algorithms implement various probabilistic, statistical and optimization techniques which allows the system to review/ learn from the past records and to identify the complex patterns/solutions from the large, complex and noisy structured semi-structured or unstructured datasets [16], [27]. Augmenting machine learning with image processing techniques facilitate automization of pap-smear analysis and generate authentic and accurate results in a faster way [4], [12], [16]. Machine learning algorithms such as Artificial Neural Network (ANN) [3], [4], [10], [28], Neural Network (NN), Support Vector Machine (SVM) [3], [14], K-nearest neighbor (KNN) [3], Linear Discriminant Analysis (LDA) [17], [21], Decision Trees [10], [17], [29], Random Forest (RF) [4], [14], [15], Gray Level Co-occurrence Matrix (GLCM) [28], Multivariate Adaptive Regression Splines (MARS), spatial fuzzy clustering algorithms, Probabilistic Neural Networks (PNNs), Genetic Algorithm, C5.0, Classification and Regression Trees (CART) and Hierarchical clustering algorithm are being used at various stages of automatic cervical cancer detection [3], [9], [30], [31], [32].

For this review, we have studied and covered 2016 onwards state-of-the-art covering 18 research papers and 20 review papers from IEEE, Springer, Elsevier, ACM, Research Gate, Hindawi, etc. Table 1 and 2 depicts the summary of review. Efficiency and accuracy of ML algorithms relies on Dataset. Most of the reviewed papers introduced a common procedure where image processing and machine learning approach is experimented [8]. Image acquisition, pre-processing, image segmentation, feature extraction, feature selection, and classification [4], [16], [17], [32]. In the first phase image/ data acquisition, two dataset were most popular which are Herlev dataset for single cell images and SIPaKMed for the multi cell Pap-smear images. Apart from these, TCT images from collaborating hospitals [2], images captured through android devices with particular resolution and microscope are being used for the processing [22]. The second phase is the pre-processing phase. The digital images collected through various modes might contain a plethora of unnecessary objects such as noise, low resolution, blurriness, etc. Hence, various image processing techniques are implemented either to correct the data or to extract adequate information. In the third phase, Image Segmentation, cervical cell images' nuclei or cytoplasm are segmented from either isolated cells, touching cells or overlapping cells [4], [10]. However, in some applications of machine learning, the segmentation phase is optional for example implementation of basic CNN [33]. The next stage emphasizes extraction and selection of necessary features such as shape, size, color intensity, textures etc. from the segmented regions that are helpful in detection of abnormal cells from the cervix images [16]. In the last phase, classification of the images into different classes such as superficial cell, intermediate cell, columnar cells, dysplastic cells, carcinoma in situ, etc. to differentiate between normal and abnormal cells are achieved [1], [11].

In the current state-of-an-art, most of the researchers emphasize single cell cervical images which are easy to work with and provide successful results. Quite a few researchers focused on using multi-cell images and generating successful detection of abnormal cells [33]. The difficulty level of identifying the malignant cells and classifying them according to their degree of infection is immense. However, many expert research analysts have achieved success in delivering results using multi-cell cervix images with high accuracy and efficiency [30].

This article is fashioned in the given manner. Section 2 presents the background of image processing techniques and machine learning concepts in the healthcare and its applications. Section 3 pdiscussed related work from the research and survey papers collected from google scholar, science direct, research gate. The key words used to collect papers are machine learning, image processing, medical image processing, cervical cell detection, pap-smear image analysis, automatic cervical cell detection and classification. Followed by section 4, that covered comparative study and discuss various approaches with pros and cons. Lastly, section 5 concludes the whole study and discussed future direction.

2 Background

2.1 Image Processing Techniques For Healthcare

Image Processing (IP) is a computer technology that process images to analyze and extricate fruitful information from them. Due to technological advancement IP techniques have become widely popular all over the world. Among all other application sector of IP, medical sector gained a lots of popularity due to its wide range of imaging tools and technologies for the internal diagnosis of body parts [20]. Figure 1 depicts the working process of IP. It is a sequential process from lower level to higher level processing. It is divided into three categories according to their outcomes and applications viz. low-level, mid-level and high-level processing [23]. Low-level image processing techniques take images for the input and produce updated, enhanced, denoised and improved

Table 1: Details of state-of-an-art Journals-Yearwise

Year	2022	2021	2020	2019	2018	2017	2016
Papers	9	5	7	6	4	3	4

Table 2: Details of state-of-an-art Journalwise

Year	IEEE	Springer	Elsevier	ACM	Research Gate	MDPI	Others
Papers	4	3	2	2	2	4	21

appearance images for the output. In most cases, low-level processing is used for pre-processing and image enhancement. In mid-level image processing, input is images and output will be attributes extracted which can be used for various applications such as image segmentation, object recognition, and description of images as mentioned [20]. Lastly, high-level image processing is used to extract important information out of images and later use them as knowledge for making sense out of it. Input for high level processing is images and outcome can be better understanding of the images for the various applications such as object detection and identification, autonomous navigation system, intrusion detection system, theft identification, and scene recognition. [34].

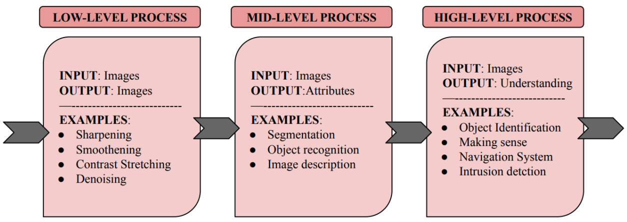

Figure 1: Image Processing Levels

Biomedical images are an integral part of medical science that demonstrates the human biological system in order to understand the nature of the human body [15], [23], [24], [31], [35]. A plenty of tools and techniques are being developed to inspect the human body and identify and analyze the diseases for the medical diagnosis [30]. X-rays [24], CT-Scan (computerized tomography scan) [24], Mammography [36], MRI (magnetic resonance imaging), ECG (electrocardiogram) [37], Ultrasound sonography [24], [38], MRA (magnetic resonance angiography) [39], stereo endoscopy, PET (positron emission tomography) [40], doppler techniques, photoacoustic imaging [41] are some of the example of it [20], [23], [24].Some of the applications of various medical image processing techniques [24], covering X-Ray to MRI is demonstrated in figure 2 [11].

2.2 Medical Diagnosis In Image Processing- A General Approach

Conventional medical imaging techniques provide prospective and guaranteed advancement in science and healthcare [23], [27]. Imaging technologies in medicine helps doctors to see the internal body parts for quick and easy diagnosis, and keyhole surgeries for unreachable body parts without cutting it open. The digital image processing involves steps mentioned in the figure 3 below which provides general insight into how medical diagnosis can be performed by degrees [23].

Step 1 starts with image acquisition which includes capturing digital images through imaging technologies such as x-rays, MRI, CT-scan etcetera. In step 2 image pre-processing and noise reduction is performed to enhance the appearance of the images and make it more readable for the further processing [30]. In the next step image segmentation, various operations such as edge detection is performed which is an essential operation in medical image analysis systems and is used to recognize human anatomy such as vessels, liver, brain, cervix, and breasts. In step 4, various algorithms are performed to successfully detect the bones, organs, nodules in lungs, tissues and cells so that detailed analysis can be performed on the selected areas in step 5 to identify the problems in the areas of interest in the human body and making sure that the organs are working normally. The analyzing algorithm of digital image processing systems focuses on measurement in terms of volume, growth, functioning of the organs, cardiac functions, checking for stroke related problems and many more. The last step consists of image classification and diagnosis related techniques which help in determining the existence of cancer, differentiating between normal and abnormal cells, identifying the actual cause and problem and suggesting primary cure treatment on an early basis.

2.3 Machine Learning For Healthcare

Machine Learning (ML) is a branch of Artificial Intelligence (AI) which has the ability to learn from the past noisy, complex and large data records through training and predict the future results for the given problems [9], [16], [26].

The ML approach for medical diagnosis applications has gained momentum in recent years [26], [42]. The applications of using machine learning algorithms for healthcare include providing personalized medications with precision, detailed analysis and examination of radiology images and data, computer assisted prognosis, clinical workflow monitoring including patients, etc. Various important aspects of applying machine learning techniques for the healthcare field are portrayed [43] in figure 4. Managing the patients' records and medical histories, suggesting primary care treatments of chronic diseases, cancer screening, surveillance, tumor characterisation, and drug discovery are the keyline areas where machine learning techniques provide satisfactory outcomes in terms of prediction and detection [27].

3 Related Work

Loe Zhe Wei et al 2020 [3] presented a review article for detection and classification methods for cervical cells in

Figure 2: Image Processing For Healthcare

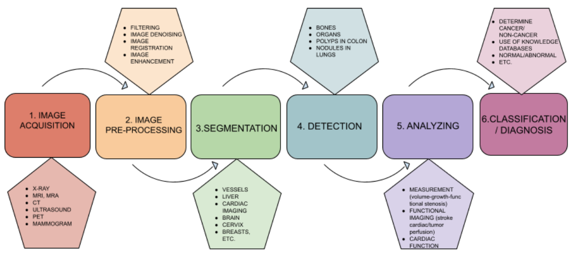

Figure 3: Medical Diagnosis in Image Procesisng- A general Flow

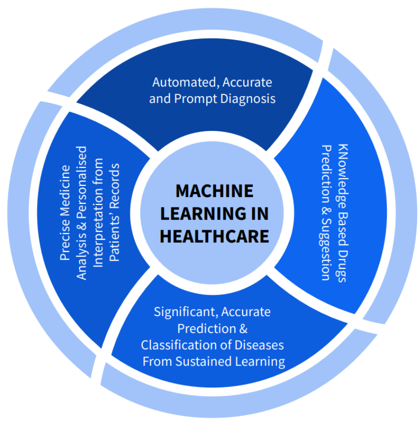

Figure 4: Machine Learning for Healthcare

automated systems. They discussed the pros and cons of each individual method. They concluded that identification of the overlapping cells of the nucleus is one of the most significant findings during the whole screening process. A comparative study provided that can help to figure out difficulty levels associated with each technique and simultaneously provides basic knowledge of generating and using their own algorithms as per the requirement.

Yessi jusman et al 2014 [17] presented a comparative study for the screening of cervical carcinoma and divided the whole process into two approaches which are cellular level and tissue level. They suggested, the automation can be done either on images or on spectra. In the feature extraction step morphology operations on texture, shape and/or intensity of the cell or tissue images are extracted whereas, for spectra, the focal features are height of intensity, shift or wave number and corrected area under peaks of the spectra. For feature selection techniques used are sequential backward, forward and floating search methods, discriminant analysis and pca. classification process can be performed using NN, SVM, KNN, LDA, and Decision Trees. They suggested that 6 different types of cervical precancerous data such as cytology, FISH, electromagnetic spectrum, cervicography, colposcopy,and HSDI can be used [7], [15], [17], [18], [20], [21]. They shown that cellular level data, namely, cytology, FISH, and electromagnetic spectrum can achieve better outcomes when compared with tissue level data such as cervicography and colposcopy.

S. PradeepKumar Kenny et al, 2016 [1] represented a comparative study between single and multiple features extraction methodologies. Segmentation applying multiscale watershed technique is explained for the feature extraction process.Data mining techniques to identify suitable features to figure out different stages of cancer are used with the 100 percentage successful detection ratio.Later, the conclusion was derived that rather than single feature, combination feature set technique is significantly better.

Wan Azani Mustafa et al 2020 [10] presented a thorough review for the cervix cancer detection based on nucleus segmentation and classification techniques.For nucleus detection, many approaches are being used to identify signs of human papilloma virus using various tools such as Matlab 2009a. The data processing includes picture collections, thresholding, noise reduction, filtration, and many more methods used to identify cells either normal or abnormal. For the classification, according to the authors, the results are controversial. The disadvantages of pap smear test led the researchers towards automated classification for abnormal cells in the cervix area.The proposed system includes feature extraction, feature selection, classification techniques such as svm, neural networks (ANN,DNN). fuzzy logic, bayesian network, KNN, decision tree etc. Pros and cons of all the algorithms are discussed. One of the more important conclusions is that it's a challenging task to detect the overlapping cells.

Wasswa William et al 2018 [16] represented a detailed review on automated cervical cancer detection. All the papers reviewed are gathered through 4 scientific databases using a set of keywords. CHAMPS software is used to perform segmentation by using algorithms KNN and SVM which are reportedly excellent classifiers for the cervical cancer detection with accuracy over 95% when they are applied for more than one class classification.

Yash Singh et al [31] provided an ordered review of cervical cancer screening algorithms. Different algorithms of segmentation and classification of cervical cancer screening are explained considering various parameters such as size of the datasets used, accuracy, drawbacks etcetera. various ML algorithms are used for early detection and classification of malignant cells/ tissues in the cervix area, SVM, GLCM, RF Trees, CART, Hierarchical clustering, c5.0, MARS, K-,eans clustering algorithms, genetic algorithms, probabilistic neural networks are such ML algorithms that are used for feature extraction, segmentation and classification operations.

Sarah L. Bedell et al [15] presented a review on new screening technologies for cervical cancer detection which have more potential to reduce the mortality ratio occurring due to cervical cancer incidents in the underdeveloped countries. description on how cancer gets generated along with what are the screening methods conventional and modern.

B. Chitra, S. S. Kumar 2021 [8] presented a review on automated screening of cervical cancer happening all around recently along with their pros and cons and their effectiveness. The emphasis is being put on recent soft computing techniques for detection of cervical cancer cells. An insight has also been provided on how soft computing techniques can be helpful in segmentation as well as classification of cervical cancer diagnosis.

Elmer Diaz et al 2020 [12] have reviewed a plethora of researches and provided a system research flow for the analysis of cervical cancer. According to their analysis there are mainly 3 things to focus on : risk factors that cause cervical cancer, precautionary measures, and techniques to detect the cancer successfully. Considerable risk factors includeHPV infection, behavior sexual, psychosocial, economic, cultural , health and reproduction . A preventive measure is Vaccine for HPV and detection techniques deep learning techniques provide improved accuracy for various operations.

Xiangu Tan et al 2021 [2] proposed a deep convolutional neural network based TCT (thin prep cytologic test) cervical screening model in order to help pathologists for the whole diagnosing process. automated deep learning algorithms attain high precision fast cancer screening.collected TCT images from the collaborating hospitals were divided into three datasets: training, validation and testing of faster R-CNN system. The proposed model was able to differentiate positive and negative cells and focused on sensitivity and specificity parameters. A very small computational cost based proposed model is likely to be a part of the foundation of preliminary medical services.

Yao Xiang et al 2020 [33] implemented an efficient automated cervical screening system using CNNbased object detection detection method using YOLOv3 as a baseline omitting segmentation phase unlike other researches carried out these days. Their model achieves 97.5% sensitivity (Sens) and 67.8% specificity (Spec) on cervical cell image-level screening. The system is also able to provide location of the malignant cell along with the cell classification.

Kyi Pyar Win et al 2020 [4] proposed a four step cervical screening structure. Cell segmentation is able to detect cancer cell nuclei using shape based approach inorder to segment overlapping cytoplasm using marker-based watersheds technique. Next, the feature extraction process helps in focusing three important features such as texture, shape and color from cytoplasms and nuclei using the GLCM. For feature selection, RF (random forest) algorithms are used to reduce the complexity of the proposed model as well as training time of the machine learning algorithms. At last, in the classification step, a bagging ensemble classifier is used by combining the outcome of the five different classifiers LD, SVM, KNN, Boosted trees, and bagged trees. The proposed system's outcomes focus on two-class and five-class classification with accuracy level 98.27% and 94.09% respectively. Benefit of the proposed system is it helps identify normal and abnormal cells and it also provides better classification results compared to individual five classifiers.

Dr. S. Athinarayanam et al 2016 [28] presented an automated cervical cancer cells classification system in order to overcome the detection errors considering thickness, overlapping of cells and other unwanted substances identified by the cytologists during pap-smear analysis. The whole proposed system consists of segmentation followed by feature extraction with SVM classification and at last results of classification components are compared with KNN and ANN techniques.

Wassawa Willian et al 2019 [44] have developed a tool for diagnosis and classification from pap-smear images. Starting from scene segmentation which is achieved via a trainable weka segmentation classifier and a sequential elimination approach for noise removal process. Later feature selection process using simulated annealing is achieved with wrapper filter followed by fuzzy c-means classification is processed.The whole process is applied on 3 varied datasets that includes single cell and multiple cell pap-smear slide images. 3 parameters accuracy, sensitivity and specificity are considered for the classification results on each individual datasets. Results generated at their end suggest that proposed method outperforms many of the existing algorithms in specificity (97.47%), sensitivity (99.28%), and accuracy (98.88%) when applied on Herlev benchmark pap-smear dataset.

Vidya Kudva et al 2018 [22] have elaborated that whether classifying image patches as normal or abnormal using shallow layer CNN and its feasibility. Input datasets include cervix images after applying acetic acid using an Android device in 102 women. Image patches extracted via different techniques are classified using a shallow layer CNN which is composed of layers of convolutional, rectified linear unit pooling, and two fully connected layers resulting in 100% accuracy.

Shrishti Gautam et al 2018 [29] suggested patch based approach for nuclei in single cell segmentation by using CNN. A CNN based transfer learning approach for classification has been suggested along with a decision tree based approach. According to the results, decision tree based classification outperforms multi-class classification with transfer learning. Their final outcome proves that accurate segmentation is not required for the classification with deep learning.

Jiayi Lu et al 2019 [45] presented a combinational approach for predicting the risk of cervical cancer successfully. To address challenges associated with cervical cancer using a voting strategy, along with data correcting mechanisms which improve performance for enhancing the robustness of prediction are proposed. As compared with others, they shown that their proposed system is more implementable and scalable.

Shanthi P B1 et al 2019 [11] proposed a CNN based approach for the detection of malignant cells and classification of the cells into appropriate stage categories of malignancy of the cell. A huge dataset was being prepared by combining the Herlev Dataset carefully and feature extraction operation was being performed by using preprocessing and segmentation techniques in the sequential manner. The CNN model used here not only generates results of feature extraction successfully and extracting features that includes shape, edges, size, colors in classification stage but also differentiate the cervical cell images into various grades (normal, mild, moderate, severe, and carcinoma) according to their degree of malignancy. Final outcome of their classification process provides the result into 3 sets that includes single cell enhanced images, contour extracted images, and binary images sequentially. The result shown accountable accuracy in classification of various degrees of cervical cell images.

4 Comparative Study of Different Techniques at Each Level

The overall understanding of how the computer assisted cervical cancer screening process works is summarized in this section. The table 3 covers set of techniques and algorithms for various stages of screening process which includes stages such as image acquisition, pre-processing, feature extraction followed by selection, segmentation followed by classification along with the additional information such as possible outcomes, limitations and future work, set of parameters, tools and technologies used by the researchers in the current reasearch.

From the comparative study shown in table 3, it is quite evident that some techniques, approaches and algorithms are more frequent in use. The study reflects that an open source datasets Herlev dataset (single cell) [1], [4], [11], [16], [28], [29], [44], [46] and SIPaKMeD (multicell) of Pap-Smear images are most widely used. However, other possible approaches to get the desired dataset are collaborating with hospitals and/or laboratories to get TCT [2] or Pap-Smear images, and capturing Cervix images after applying 3%-5% acetic acid using android device with specified resolution or converting the slide data into digital form by means of some digital medium such as camera [22].

For preprocessing and image enhancement most used approaches are noise removal techniques that includes various filters [42] such as median filter [16], [29], gaussian filters [16], sobel vertical-horizontal filter [11], [16], and some smoothing filters, contrast enhancement- CLAHE (contrast limited adaptive histogram equalization) technique [4], [8], [17], [21], [27]. For feature extraction , techniques used are ANN [3], [4], [10], [27], [28], [44], CNN [2], [3], [31], GLCM [4], [28], [31], for the extraction of visual features, texture features, shape features etc [44]. Feature selection can be implemented using faster R-CNN [2] for the features such as cytoplasm, nuclear shape, fluid color, Random Forest algorithm [4], [14], [15], SVM [4], [10], [14], [16], [17], [28], [31], LBP [22], DNN [11], [27], [47].

Many of the researchers have incorporated segmentation phase while others have simply omitted by claiming that after the feature selection process, classification algorithms can be applied directly in order to achieve more accurate results. Segmentation techniques used are shape based iterative methods, marker-controlled watershed approach to segment overlapping cytoplasm [4], Modified Otsu Thresholding Algorithm [28], [44] edge detection using sobel filter, Hessian matrix, Gabor filter for boundary detection(canny edge detection, texture filtering using mean, variance, median, maximum, minimum and entropy filters [1], Random subset feature selection (RSFS) [14], [22], and Patch based CNN based approach [2], [29]. The classification is implemented using machine learning algorithms that includes Bagging ensemble classifier (LD, SVM, KNN, BOOSTED TREES, BAGGED TREES) [4], [7], [32], SVM [3], [9], [14], [28], [45], ANN [3], [4], [10], [28], KNN [3], [4], [10], [14], [16], [17], [28], Fuzzy c- means algorithm [3], [44], Deep learning methods using transfer learning [19], [30] on Alexnet on both segmented and non-segmented single [29], Logistic regression [45], MLP (multilayer perceptron) [45], Decision Tree classifier [9], [14], [45], and Convolutional Neural Network (CNN) [2], [3], [33].

The automated cervical cancer screening system has the ability to classify between normal and abnormal cells from both single cell and multicell dataset according to the classification algorithms. The outcomes are mainly measured using parameters which are accuracy, sensitivity and specificity and generates outcomes in multiclass classification. From the results shown in the summary table 3 on existing literature, Bagging ensemble classifier is able to provide accuracy in two class and 5-class classification [4] 98.27% and 94.09% sequentially, SVM provides 86% accuracy [28], Fuzzy c-means classifiers single cell gives accuracy 98.88%, sensitivity-99.28%, specificity-97.47%, Multi-cell provides accuracy-97.64%, sensitivity-98.08%, specificity-97.16%. Pap-smear data generates accuracy-95%, sensitivity-100%, specificity-90% [44], 100% accuracy is achieved using shallow CNN [22], Deep learning methods using transfer learning on Alexnet are able to provide accuracy (2-class:99.3%, 7-class: 93.75%) [29].

5 State-of-the art: Cervix cancer detection and classification

5.1 Automated Cervical Cancer Detection Workflow

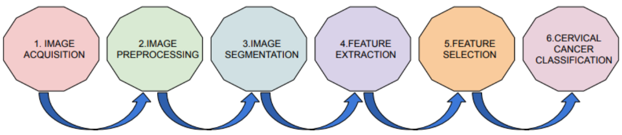

Figure 5: Cervical Cancer Detection Workflow

From the current literature, it is observed that cervical cancer detection and classification follows six main steps as shown inthe figure 5. It provides a general approach for the whole screening process which consists of process steps mentioned here. Each and every stage is equally important. Hence, their detail study is essential and so been explained in detail in the following subsections.

5.2 Invasive Cervical Cancer Diagnosis- Available Screening Methods

Squamous cell cancer is a dominant cause of cancer deaths in women in the growing world. Almost every day new technologies are being developed, implemented and tested for fast, efficient and cost-effective cervical cancer screening and medical treatment [15], [43]. Screening a woman for HPV (human papillomavirus) and cervical Dysplasia can considerably reduce the risk of cervical cancer deaths [7]. The Diagram [6] showcases available cervical cancer screening imaging technologies. The Papanicolaou test being manual cervical screening process and used for detecting precancerous changes in cervix cells using shape and color like features of the cervix cell nuclei and cytoplasm regions [4]. Samples collected through pap-smear tests are observed under microscope to find out atypical development of cells which leads to precancerous changes. However, this is a very time consuming, and laborious analysis technique. The cervical cancer diagnostic screening methods have two approaches, cell-level and tissue-level [8], [17]. The cellular-level approach includes pap-smear, Human papillomaviruses- Deoxyribonucleic acid (HPV-DNA) testing, liquid based cytology (LBC) [7], [46], and electromagnetic spectroscopies [8]. The tissue-level screening technique includes Visual inspection with Lugol's iodine (VILI) or Visual inspection of the cervix with acetic acid (VIA) [19], [21]which is visual inspection of tissues after applying Lugol's Iodine or Acetic Acid, colposcopy, cervicography, and hyperspectral diagnostic imaging (HSDI) [7], [15], [17], [22], digital cervigrams, mobile phone images, pocket colposcopes. However, each of the techniques have their advantages and disadvantages and all the techniques mentioned highly skilled experts for the judgment or prediction of the results. There are various advanced technologies available for automated screening such as AutoPap 300, Focal Point, and Thin-Prep Imaging systems (TIS) approved by United States Food and Drug Administration (USFDA) [17], [26].

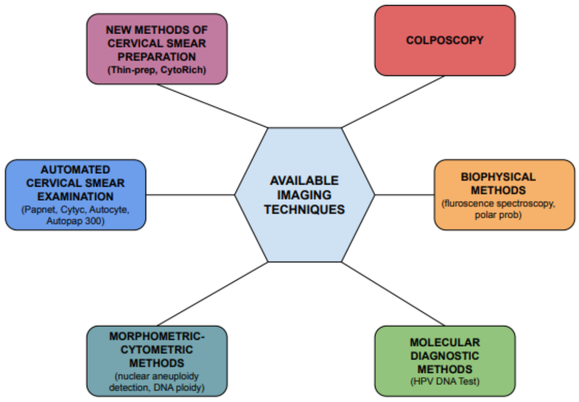

Figure 6: Invasive Cervical Cancer Diagnosis- Available Screening Methods

5.3 Image Acquisition

Image acquisition in IP is a process of acquiring images from the verified sources/ techniques which can be further processed. Many researchers are using available open source datasets for cervical cancer cells called Herlev (Single Cell) and SIPaKMed (Multi Cell). Also, other dataset sources such as AINDRA [29], HEMLBC (H and E-Stained Manual Liquid Based Cytology) [7], TCT images from collaborating hospitals and laboratories [2], [4], ISBI (International Symposium on Medical Cytology), NCI (National Cancer Institute) dataset can be acquired for primary research.

5.4 Image Preprocessing and Enhancement

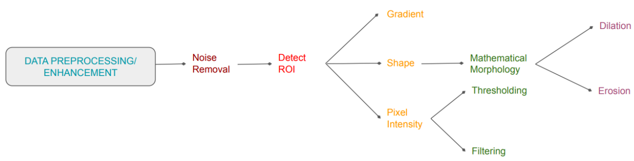

Figure 7: Image Pre-Processing And Enhancement

Being an utmost important step in computer assisted cervical cancer screening procedure, Image Enhancement provides fruitful outcomes for preprocessing and further process [6]. In the preprocessing, various operations such as increasing and/or decreasing contrast, smoothening, sharpening, removal of noise, filtering are applied to improve the images and make them suitable for next operation/procedure [8], [21]. Various noise removal filters such as mean, median, sum of squares, Gaussian filter etc. are being used for preprocessing and histogram or contrast stretching algorithms such as CLAHE is widely used for image enhancement [4], [8], [17], [21], [27]. Figure 7 depicts the procedure covering various techniques to identify features like gradient, shape and pixel intensity, radius which can be identified using techniques such as mathematical morphology (dilation, erosion, opening, closing) [8], filtering and thresholding methods [10] and making them suitable.

5.5 Image Segmentation

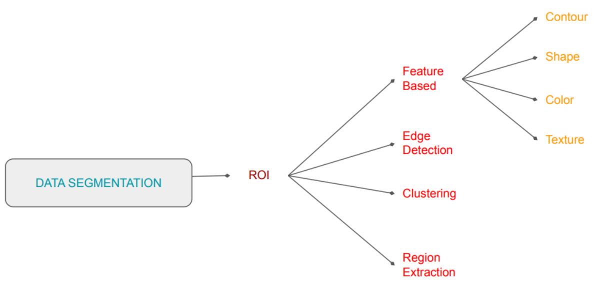

Figure 8: Image Segmentaion

In step 3 of automated screening process, the outcomes of step 2 i.e. enhanced and preprocessed images are further processed to segment the regions of interests (ROI) of cells [21], [46]. The ROI can be segmented using variation of techniques which can be features like contour, shape, color or texture based, edge detection based, clustering based , or region to be extracted based [8]. Cell segmentation aims at nuclei or cytoplasm of the cells out of which segmenting the nuclei is easier as compared to later. Both nuclei and cytoplasms are used for overlapping cell segmentation [4], [32] . Isolated cells, overlapping cells, and touching cells segmentation are performed where shape based iteration methods considering features such as area, intensity value, solidity, major and minor axis length for nuclei and watershed transform approach for cytoplasm where smoothening of boundaries is performed using edge smoothing methods [4].

5.6 Feature Extraction

After the image segmentation, its output passes to features extraction process. In this stage, various image features such as texture (rough texture for abnormal nucleus), shape (smooth, circular and oval boundary specifies normal nucleus), ratio, color intensity, chromaticity (cancerous nucleuses are darker in shade), size (radius, area, perimeter of the cell) etcetera are extracted. Various factors as shown in the figure 9 are considered for both cell and tissue feature extraction process [8]. Many algorithms are available to extract features. For example, texture features can be extracted by applying co-occurrence matrix, wavelet technique, mathematical morphological operations, clustering techniques, thresholding approach and many more [8], [10], [32]. Differentiation between normal and abnormal cells are performed using color and shape features for which various color models such as RGB , HSV, gray-level histogram, watershed technique etc are used [4], [8], [17].

5.7 Feature Selection

In the process of automated cervical screening, the next step is Feature Selection used for enhancing the performance of the classifiers. Selection of appropriate features

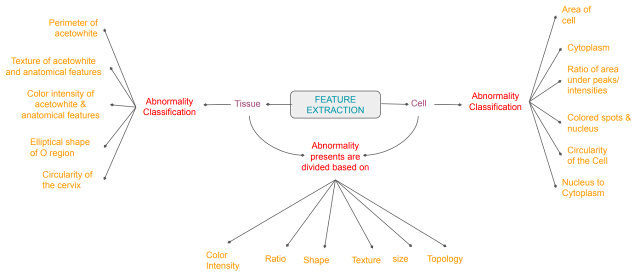

Figure 9: Feature Extraction

plays a vital role as it helps reducing size of dataset and at the same time transforms high-dimensional input to low-dimension input form. From the thorough review [8], Principal Component Analysis (PCA), and Discriminant Analysis (DA) are identified as popular selection and extraction algorithms [14]. The PCA technique [6] uses the principle of orthogonal linear transformation to convert data to its new co-ordinator. It is one type of linear classifier used for Pap-Smear images classification mostly. On the contrary, DA technique creates a new value known as the discriminant function score. The DA method is similar to computing Eigenvalues. Both the PCA and DA techniques are used to select features such as Texture , Shape, Ripplet Description. Texture Features can be classified into energy, skewness, mean, energy, variance, contrast, sum of homogeneous features, cluster, sum of squares/ average/entropy/variance/energy, etcetera [1]. Shape Features are further classified into eccentricity, compactness, circularity, area and perimeter. Ripplet Description can be classified into color and texture which are used to identify ripplet descriptors [8].

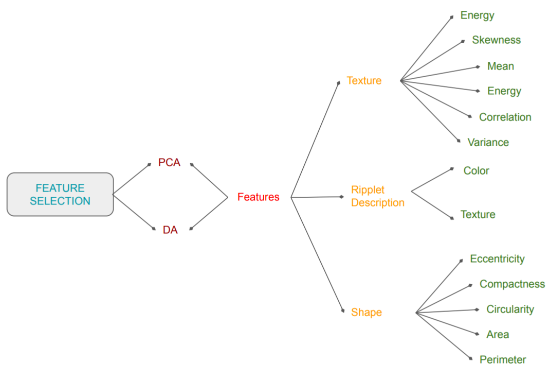

Figure 10: Feature Selection

5.8 Cervical Cell Screening Algorithms

Figure 11 showcases the summary of algorithms used so far for an automated cervical cancer screening system [43]. The algorithms and techniques summarizes the methods of various research survey papers reviewed based on database size, accuracy, drawbacks, etc. [31], [48] Various machine learning algorithms [6], [8], [31], [32], [49] widely used are Decision Trees [10], [14], [17], [29], Support Vector Machines (SVM) [3], [14], C5.0 classification model, Random Forest Trees (RF Trees) [4], [14], [15], [35], Multivariate Adaptive Regression Splines (MARS), Hierarchical Clustering, Gray Level Co-Occurrence Matrix

(GLCM), Classification And Regression Trees (CART), K-means Clustering Algorithm [4], Probabilistic Neural Networks (PNN), Genetic Algorithms, Convolutional Neural Networks (CNN) [2], [3], [49], Artificial Neural Networks (ANN) [3], [4], [10], [28], Deep Neural Networks (DNN) [47], K-Nearest Neighbors (KNN) [3], [10], [14], [16], [17], [28], Bayesian Network [35].

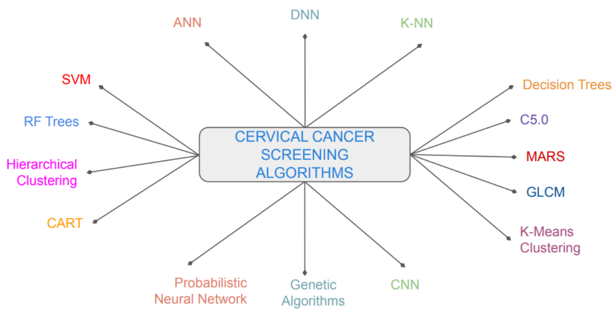

Figure 11: Cervical Cancer Screening Algorithms

5.9 Cervical Cell Image Object Categories

Cervical cell/ nuclei can be categorized into various categories mentioned in the figure 12 i.e either normal and/or abnormal (cancerous benign and/or malignant) cell/nuclei [4], [21], [33]. There are typically 10 different object categories for cervical cells which are normal, Atypical Squamous cells-Undetermined Significance (ASC-US), Atypical Squamous Cells- cannot exclude HSIL (ASC-H), Low-Grade Squamous Intraepithelial Lesion (LSIL), High-Grade Squamous Intraepithelial Lesion (HSIL), Atypical Glandular Cells (AGC), Adenocarcinoma (ADE), Vaginalis Trichomoniasis (VAG), Monilia (MON) and Dysbacteriosis (DYS) [2], [21], [33]. As per the current literature, cervical cell images can be categorized into two categories Normal (Immediate, Columnar, Superficial) and abnormal (Moderate Dysplasia, Light Dysplasia, Severe Dysplasia, Carcinoma in Situ) for single cell images and three categories like Normal (Superficial-Intermediate, Parabasal), benign (Metaplastic) and abnormal (Dyskeratotic, Koilocytotic) for multi-cells images [4], [16].

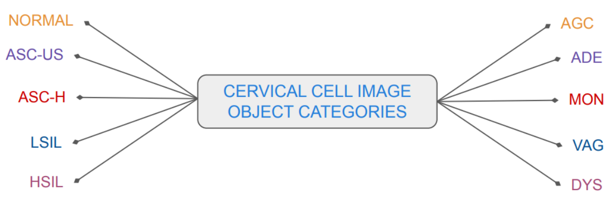

Figure 12: Cervical Cell Image Object Categories

6 Conclusion and Future work

For this literature survey, various machine learning papers and articles have been reviewed. In this review, around 50 papers from various journals and conferences are analyzed which are carefully selected through decorated publications which include IEEE, Springer, Elsevier, Sage, Acm and others from the year 2016 to 2022. Most of the papers have mentioned the usage of Herlev Pap-smear Dataset available open-source for single cell multi-class classification. This review should help the novice researchers the primary assistance in developing different approaches to improve the existing outcomes or generate new algorithms. The Neural network based approaches are widely popular for cervical cell image classification. Most widely used are SVM, CNN and KNN. Various Network models such as VGG-16, Alexnet, ImageNet can be incorporated to improve the efficiency of the output. As the available open source datasets are limited in terms of the size, more data is required to improve the accuracy of the results from the collaborating hospitals, laboratories and private organizations. Also, image augmentation techniques can be experimented to increase the size of the training datasets which results in improved accuracy in the test datasets. Applying appropriate preprocessing techniques can help in improving the results of segmentation and classification algorithms.

Table 3: A Comparative Study Of Different Techniques At Each Level

PAPER	IMAGE ACQUISITION	TOTAL NO. OF IMAGES	DATA PREPROCESSING	FEATURES EXRACTION	FEATURES SELECTION	SEGMENTATION	CLASSIFICATION	OUTCOMES	LIMITATIONS/ FUTURE WORK	PARAMETERS	TOOLS AND TECHNOLOGIES
Xiangu Tan et al. (2021) [2]	TCT images from collaborating hospitals	Training data set(13775), validation dataset(2301), test dataset (408030 from 290 scanned copies)	Seamless slider technology, lesion labelling	ANN, CNN	Faster R-CNN to extract Cytoplasm, nuclear shape, fluid base color	-	-	Separated normal and abnormal images, minimize the rate of missed diagnosis, improved speed and accuracy,	Difficult to differentiate between malignant and normal cells due to high no. of overlapping cells.hard to find characteristics of cells,dataset does not have enough samples of TCT images	Sensitivity, Specificity considering TN,TP,FN,FP,PPV,NPV	The labelling software: Labellmg(Version 1.5.0),
Kyi Pyar Win et al (2020) [4]	Pap smear images. SIPaKMeD (multi-cells), Herlev (single cell)	Herlev dataset (917 images), SIPaKMeD dataset (966 images, 4049 cells)	Image enhancement(noise removal- median filter, contrast enhancement-CLAHE(contrast limited adaptive histogram equalization))	3 features extracted-shape, size,texture and color features using GLCM (gray level co-occurrence matrix)	Random forest algorithm	Segmentation of nuclei regions and cytoplasm regions Nuclei detection-shape based iterative method, overlapping cytoplasm separation-marker-controlled watershed approach	Bagging ensemble classifier (LD, SVM, KNN, BOOSTED TREES, BAGGED TREES)	98.27% accuracy in two class classification, 94.09% accuracy in 5-class classification using SIPaKMed dataset	Use of other classifiers can further enhance the result	Accuracy, sensitivity/recall, specificity, precision and F_Measure	-NA-
Dr. S. Athinarayanam et al.(2017) [28]	Pap smear images: captured with resolution of 0.201 μm/pixel from public database of cervical cancer,Herlev university hospital	Normal (40 images), abnormal (60)	Gray scale conversion,Adaptive Path Smooth Filter to remove small noise and preserve sharp edges	Texture feature extraction-GLCM	energy, correlation, entropy, contrast and homogeneity using GLCM	Modified Otsu Thresholding Algorithm (background removal,cytoplasm and nucleus detection)	SVM, ANN, KNN	The proposed method is effective for the classification of Pap Smear Cell image into normal and abnormal compared the output of SVM with ANN and KNN and achieved 86% accuracy.	-NA-	Sensitivity, accuracy, specificity	-NA-
Wassawa Willian et al.(2019) [16]	Pap-smear images: Herlev benchmark pap-smear dataset	Dataset1: 917 single cell of Herlev Dataset 2:497 full-slide images Dataset 3: 60 images from Mbarara Regional Referral hospital (MRRH)	Three phase elimination scheme :debris removal, image enhancement: A contrast local adaptive histogram equalization,	Dataset divided into 7 class considering size, area, shape, and brightness of nucleus and cytoplasm. 3 geometric features (solidity, compactness and eccentricity) and 6 texture features (mean, standard deviation, variance, smoothness, energy, and entropy)	Simulated annealing integrated with a wrapper filter. The performance of the feature selection is evaluated using a fitness value evaluated using k-fold cross validation.	Scene segmentation: trainable weka segmentation (TWS). Noise reduction, edge detection using sobel filter, Hessian matrix, Gabor filter for boundary detection(canny edge detection, texture filtering using mean, variance, median, maximum, minimum and entropy filters.	Fuzzy c- means algorithm	Developed their own tool PAT. single cell :accuracy-98.88%, sensitivity-99.28%, specificity-97.47%. Multi-cell: accuracy-97.64%, sensitivity-98.08%, specificity-97.16%. pap-smear : accuracy-95%,sensitivity-100%, specificity-90%. FN, FP, classification error rate are 0.00%, 10.00%, 5.00% sequentially.	Haven't included cervical cancer risk factors assessment into the tool.	Accuracy, sensitivity, specificity for single cell and multi cell,and pap smear images.	Trainable Weka segmentation tool , MATLAB, Java Graphical user interface(GUI)

Continue on the next page

10

Table 3: A Comparative Study Of Different Techniques At Each Level (cont.).

PAPER	IMAGE ACQUISITION	TOTAL NO. OF IMAGES	DATA PREPROCESSING	FEATURES EXRACTION	FEATURES SELECTION	SEGMENTATION	CLASSIFICATION	OUTCOMES	LIMITATIONS/ FUTURE WORK	PARAMETERS	TOOLS AND TECHNOLOGIES
Vidya Kudva et. al. (2018) [22]	Cervix images after applying 3%-5% acetic acid using Android Device with 13MP camera	Total 102 out of which 42 (VIA-positive-pathologic) 60 (VIA-negative-healthy controls)	684 image patches of 15*15 pixels manually extracted (275-extracted from expert annotated AW regions-positie examples)(409-from non-AW regions- negative examples)	Randomly select weights for filter, pass training set through the n/w, update filter weights,compute cross entropy error function, repeat the same	Traditional approach for comparison: SVM- extract features such as color, haralick, local binary pattern(LBP)	Traditional approach for comparison: Random subset feature selection (RSFS)	CNN	100% accuracy is achieved using shallow CNN	Training CNN requires complex computations.	Accuracy, training loss,	Intel processor Celeron and 4GB RAM, MATLAB R2017a. Python programming
Shrishti Gautam et al. (2018) [29]	Pap-smear (single cell and multi-cell) dataset.Herlev and Aindra dataset	Herlev (917-single nuclei)Aindra dataset (80 multi-cell),from Bangalore, India	Median filter- noise removal.	Detection of nuclei CLAHE + thresholding	Feature based cell separation method	Patch based CNN based approach.	Deep learning methods using transfer learning on Alexnet on both segmented and non-segmented single cell images. Combination of decision-tree based classification with transfer learning which then applied to multi-cell images.	Accurate segmentation is not necessary for classification with deep learning.herlev dataset state-of-art accuracy (2-class:99.3%, 7-class: 93.75%)	-NA-	accuracy	AlexNet
Azian Azamini Abdullah et al. (2019) [18]	13 from HUSM Kubang Kerian , 102 from pap-smear	115 images	Blue channel extraction template: unwanted background is filtered out. Contrast enhancement: nucleus appears in simulated images	Initial size, boundary condition, pixel value, control template, feedback template, threshold value	-NA-	-NA-	-NA-	Proposed CNN algo. Can detect the cervix cancer cells automatically with accuracy >80%	More images to be simulated for more accurate and precise results.	accuracy	Matlab based CNN simulator.CANDY software (vismouse). Few templates such as blue channel extraction template, contrast enhancement, median filter , binary edge detection, hollow-concave.
J. Lu, E. Song, A.Ghoneim et al. (2020) [45]	Gene sequencing dataset (3000 genetic loci, belonging to 23 pairs of chromosomes)	Private dataset: 472 questionnaires (50 attributes) obtained from a chinese hospital, UCI dataset: 858 samples (32 attributes)	Data correction: Random Forest algorithm. Fill missing data :correction mechanism carefully designed using logic	History of drinking, age of first pregnancy, cervical surgery, etc.	-NA-	-NA-	Logistic regression, SVM, KNN, MLP (multilayer perceptron), Decision Tree classifier	As compared to existing methods, the proposed system is more implementable and scalable.	It doesn't have enough experimental support. Also further investigations can focus on more distinctive data that includes colposcopy images.	Accuracy, recall, precision, F1 score	-NA-

Continue on the next page

11

Table 3: A Comparative Study Of Different Techniques At Each Level (cont.).

PAPER	IMAGE ACQUISITION	TOTAL NO. OF IMAGES	DATA PREPROCESSING	FEATURES EXRACTION	FEATURES SELECTION	SEGMENTATION	CLASSIFICATION	OUTCOMES	LIMITATIONS/ FUTURE WORK	PARAMETERS	TOOLS AND TECHNOLOGIES
Shanthi P B, Faraz Faruqi et al.(2019) [27]	Pap Smear database developed by Herlev University Hospital collected using a digital camera and microscope under a resolution of 0.201 μ m/pixel	917 images distributed unequally on 7 different classes. Proposed system: 749 images with 5 class,	Data augmentation to expand the dataset for training., 5 different algorithms for Image Enhancement (Bi-Histogram Equalization with adaptive sigmoidal function combined with Sobel filter-vertical and horizontal,dynamic Fuzzy Histogram Equalization, color image processing using YCbCr color space, Fuzzy image Mapping, Genetic algorithm	Extraction of visual features such as edges, size, shape and colors out of nucleus and cytoplasm.	Shape and size of the nucleus and cytoplasm of the binarized images using deep neural network architecture	Canny edge detector to extract edge information from single cell images	Deep prediction model using CNN network to classify grades of cancer and extracting features to do so.	Cancer grade classification: normal, mild, moderate, severe, carcinoma. Outcome for 3 different sets: single-cell, contour extracted, binary images sequentially with noticeable accuracy for various class problems.	Increase in no. of classes in binarized images leads to increasing the complexity of classification, hence model requires more features for improving the accuracy	Accuracy	Used own CNN model
Yao Xiang et al. (2021) [8]	own dataset captured by digital camera Ximea MC124CG-SY-UB with 12 million pixels situated on the microscope Olympus BX40 with 20 objective. Each pixel has a size of 3.45 mm2	12,909 cervical images with 58,995 ground truth boxes and contains 10 categories objects from cervical cell images. 1014 cervical cell images with size of 4000 3000, which are consisted of 728 abnormal cell images (positive samples) and 286 normal cell images (negative samples).	-NA-	method extract high-level features automatically and detect cervical cells directly using CNN model based YOLOv3	-NA-	-NA-	YOLOv3 as a base model to detect 10 categories and then cascading a further hard example classifier to refine the 4 categories: ASC-US, ASC- H, LSIL, HSIL.	automatically detect cervical cells directly on multi-cell images. more efficient as we can extract features automatically without manual intervention and careful design for all stages. cervical cell image-level classification along with detailed location and category information of abnormal cells	-NA-	evaluation metrics used by the PASCAL VOC object detection challenge, which are average precision (AP) and mean average precision (mAP).accuracy (Acc), sensitivity (Sens) and specificity (Spec)	Darknet-53, YOLOv3, Imagenet, NVIDIA GTX 1080 Ti GPU.
Sudhir Sornapudi, et al. (2019) [46]	cytology slide data. two datasets. first set comprises 25 cervical liquid-based cytology slides provided by Becton-Dickinson (BD) Corporation using their Sure Path technique. Hamamatsu NanoZoomer 2.0-HT whole-slide scanner, Herlev Pap Smear datase	BD Corp. Data, 0.228x0.228, NDPI file type, 25 images. Herlev Data, Single cervical cell BMP file type, 917	detecting ROI bounding box coordinates.	image registration through feature based image alignment.	-NA-	-NA-	CNN models for successful classification of cytology image data, graph-based cell detection technique	Proposed approach considers realistic conditions of overlapping cells and automatically generated cleaner labeled patch image data for training and testing convolution	-NA-	Accuracy, precision, recall, f1 score, MCC	CNN models- Resnet-50, VGG-19, Densenet-121, Inception_v3

12

References

[1] S. P. K. Kenny and S. Victor, “A comparative analysis of single and combination feature extraction techniques for detecting cervical cancer lesions,” *ictact Journal on Image and Video Processing*, vol. 6, no. 3, pp. 1167–73, 2016.

[2] X. Tan, K. Li, J. Zhang, W. Wang, B. Wu, J. Wu, X. Li, and X. Huang, “Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study,” *Cancer Cell International*, vol. 21, no. 1, pp. 1–10, 2021.

[3] L. Z. Wei, W. A. Mustafa, M. A. Jamlos, S. Z. S. Idrus, and M. H. Sahabudin, “Cervical cancer classification using image processing approach: A review,” in *IOP Conference Series: Materials Science and Engineering*, vol. 917, no. 1. IOP Publishing, 2020, p. 012068.

[4] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. Myo Aung, “Computer-assisted screening for cervical cancer using digital image processing of pap smear images,” *Applied Sciences*, vol. 10, no. 5, p. 1800, 2020.

[5] W. H. Organization *et al.*, “Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation,” 2021.

[6] N. Al Mudawi and A. Alazeb, “A model for predicting cervical cancer using machine learning algorithms,” *Sensors*, vol. 22, no. 11, p. 4132, 2022.

[7] K. Kaushik, A. Bhardwaj, S. Bharany, N. Alsharabi, A. U. Rehman, E. T. Eldin, and N. A. Ghamry, “A machine learning-based framework for the prediction of cervical cancer risk in women,” *Sustainability*, vol. 14, no. 19, p. 11947, 2022.

[8] B. Chitra and S. Kumar, “Recent advancement in cervical cancer diagnosis for automated screening: a detailed review,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–19, 2021.

[9] U. K. Lilhore, M. Poongodi, A. Kaur, S. Simaiya, A. D. Algarni, H. Elmannai, V. Vijayakumar, G. B. Tunze, and M. Hamdi, “Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques,” *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022.

[10] W. A. Mustafa, A. Halim, M. A. Jamlos, and S. Z. S. Idrus, “A review: Pap smear analysis based on image processing approach,” in *Journal of Physics: Conference Series*, vol. 1529, no. 2. IOP Publishing, 2020, p. 022080.

[11] P. Shanthi, F. Faruqi, K. Hareesha, and R. Kudva, “Deep convolution neural network for malignancy detection and classification in microscopic uterine cervix cell images,” *Asian Pacific journal of cancer prevention: APJCP*, vol. 20, no. 11, p. 3447, 2019.

[12] E. Diaz, A. Ccopa, and L. Wong, “Cervical cancer: Machine learning techniques for detection, risk factors and prevention measures,” 2020.

[13] A. Reichheld, P. K. Mukherjee, S. M. Rahman, K. V. David, and R. A. Pricilla, “Prevalence of cervical cancer screening and awareness among women in an urban community in south india—a cross sectional study,” *Annals of global health*, vol. 86, no. 1, 2020.

[14] D. Y. Coşar Soğukkuyu and O. Ata, “Diagnosing cervical cancer using machine learning methods,” 2022.

[15] S. L. Bedell, L. S. Goldstein, A. R. Goldstein, and A. T. Goldstein, “Cervical cancer screening: past, present, and future,” *Sexual medicine reviews*, vol. 8, no. 1, pp. 28–37, 2020.

[16] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, “A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images,” *Computer methods and programs in biomedicine*, vol. 164, pp. 15–22, 2018.

[17] Y. Jusman, S. C. Ng, and N. A. Abu Osman, “Intelligent screening systems for cervical cancer,” *The Scientific World Journal*, vol. 2014, 2014.

[18] A. A. Abdullah, A. Fonetta, D. Giong, N. Adilah, and H. Zahri, “Cervical cancer detection method using an improved cellular neural network (cnn) algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp. 210–218, 2019.

[19] K. T. Desai, B. Befano, Z. Xue, H. Kelly, N. G. Campos, D. Egemen, J. C. Gage, A.-C. Rodriguez, V. Sahasrabuddhe, D. Levitz *et al.*, “The development of “automated visual evaluation” for cervical cancer screening: The promise and challenges in adapting deep-learning for clinical testing,” *International Journal of Cancer*, vol. 150, no. 5, pp. 741–752, 2022.

[20] G. Dougherty, *Digital image processing for medical applications.* Cambridge University Press, 2009.

[21] L. W. Habtemariam, E. T. Zewde, and G. L. Simegn, “Cervix type and cervical cancer classification system using deep learning techniques,” *Medical Devices (Auckland, NZ)*, vol. 15, p. 163, 2022.

[22] V. Kudva, K. Prasad, and S. Guruvare, “Automation of detection of cervical cancer using convolutional neural networks,” *Critical Reviews™ in Biomedical Engineering*, vol. 46, no. 2, 2018.

[23] N. Goel, A. Yadav, and B. M. Singh, “Medical image processing: a review,” *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*, pp. 57–62, 2016.

[24] S. Naseera, G. Rajini, B. Venkateswarlu, and M. Priyadarisini, “A review on image processing applications in medical field,” *Research Journal of Pharmacy and Technology*, vol. 10, no. 10, pp. 3456–3460, 2017.

[25] K. Rao and V. Rao, “Medical imaging.”

[26] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *Radiographics*, vol. 37, no. 2, p. 505, 2017.

[27] M. Y. Shaheen, “Adoption of machine learning for medical diagnosis,” *ScienceOpen Preprints*, 2021.

[28] S. Athinarayanan and M. Srinath, “Classification of cervical cancer cells in pap smear screening test,” *ICTACT Journal on Image and Video Processing*, vol. 6, no. 4, pp. 1234–8, 2016.

[29] S. Gautam, N. Jith, A. K. Sao, A. Bhavsar, A. Natarajan *et al.*, “Considerations for a pap smear image analysis system with cnn features,” *arXiv preprint arXiv:1806.09025*, 2018.

[30] J. Manhas, R. K. Gupta, and P. P. Roy, “A review on automated cancer detection in medical images using machine learning and deep learning based computational techniques: Challenges and opportunities,” *Archives of Computational Methods in Engineering*, pp. 1–41, 2021.

[31] Y. Singh, D. Srivastava, P. Chandranand, D. Singh *et al.*, “Algorithms for screening of cervical cancer: A chronological review,” *arXiv preprint arXiv:1811.00849*, 2018.

[32] K. P. Win, Y. Kitjaidure, M. P. Paing, and K. Hamamoto, “Cervical cancer detection and classification from pap smear images,” in *Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing*, 2019, pp. 47–54.

[33] Y. Xiang, W. Sun, C. Pan, M. Yan, Z. Yin, and Y. Liang, “A novel automation-assisted cervical cancer reading method based on convolutional neural network,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 611–623, 2020.

[34] R. Gonzalez, R. Gonzalez, and P. Wintz, *Digital Image Processing*, ser. Applied mathematics and computation. Addison-Wesley, 1987. [Online]. Available: https://books.google.co.in/books?id=QvBRAAAAMAAJ

[35] X. Li and X. Xie, “A disease diagnosis method based on machine learning,” in *Proceedings of the 3rd International Conference on Intelligent Information Processing*, 2018, pp. 184–189.

[36] MayoclinicStaff, “Mammogram,” https://www.mayoclinic.org/tests-procedures/mammogram/about/pac-20384806, April 2022, © 1998-2022 Mayo Foundation for Medical Education and Research (MFMER). All rights reserved.

[37] ——, “Electrocardiogram (ecg or ekg),” https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983?p=1, March 2022, © 1998-2022 Mayo Foundation for Medical Education and Research (MFMER). All rights reserved.

[38] R.-T. R. S. of North America (RSNA) or the American College of Radiology (ACR), “Ultrasound sonograpgy,” https://www.radiologyinfo.org/en/ultrasound, copyright © 2022 Radiological Society of North America, Inc. (RSNA).

[39] RadiologyInfo, “Magnetic resonance angiography (mra),” https://www.radiologyinfo.org/en/info/angiomr, March 2022, copyright ® 2022 Radiological Society of North America, Inc.

[40] J. H. Medicine, "Positron emission tomography (pet)," https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983?p=1, August 2022, copyright © 2022 The Johns Hopkins University, The Johns Hopkins Hospital, and Johns Hopkins Health System. All rights reserved.

[41] A. Neprokin, C. Broadway, T. Myllylä, A. Bykov, and I. Meglinski, "Photoacoustic imaging in biomedicine and life sciences," *Life*, vol. 12, no. 4, p. 588, 2022.

[42] J. J. Tanimu, M. Hamada, M. Hassan, H. Kakudi, and J. O. Abiodun, "A machine learning method for classification of cervical cancer," *Electronics*, vol. 11, no. 3, p. 463, 2022.

[43] S. K. Singh and A. Goyal, "Performance analysis of machine learning algorithms for cervical cancer detection," in *Research Anthology on Medical Informatics in Breast and Cervical Cancer*. IGI Global, 2023, pp. 347–370.

[44] 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.

[45] J. Lu, E. Song, A. Ghoneim, and M. Alrashoud, "Machine learning for assisting cervical cancer diagnosis: An ensemble approach," *Future Generation Computer Systems*, vol. 106, pp. 199–205, 2020.

[46] S. Sornapudi, G. T. Brown, Z. Xue, R. Long, L. Allen, and S. Antani, "Comparing deep learning models for multi-cell classification in liquid-based cervical cytology image," in *AMIA Annual Symposium Proceedings*, vol. 2019. American Medical Informatics Association, 2019, p. 820.

[47] P. Xue, J. Wang, D. Qin, H. Yan, Y. Qu, S. Seery, Y. Jiang, and Y. Qiao, "Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis," *NPJ digital medicine*, vol. 5, no. 1, pp. 1–15, 2022.

[48] H. Karani, A. Gangurde, G. Dhumal, W. Gautam, S. Hiran, and A. Marathe, "Comparison of performance of machine learning algorithms for cervical cancer classification," in *2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE, 2022, pp. 1–7.

[49] M. Kruczkowski, A. Drabik-Kruczkowska, A. Marciniak, M. Tarczewska, M. Kosowska, and M. Szczerska, "Predictions of cervical cancer identification by photonic method combined with machine learning," *Scientific Reports*, vol. 12, no. 1, pp. 1–11, 2022.