Expert Systems With Applications 141 (2020) 112951

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A fully-automated deep learning pipeline for cervical cancer classification

Zaid Alyafeai, Lahouari Ghouti*

Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Article history:
Received 25 March 2019
Revised 28 July 2019
Accepted 12 September 2019
Available online 19 September 2019

Keywords:
Cervical cancer
Deep learning
Cervix detection
Cervical region-of-interest
Convolutional neural networks
Guanacaste and Intel&mobileodt cervigram datasets

ABSTRACT

Cervical cancer ranks the fourth most common cancer among females worldwide with roughly 528,000 new cases yearly. Around 85% of the new cases occurred in less-developed countries. In these countries, the high fatality rate is mainly attributed to the lack of skilled medical staff and appropriate medical pre-screening procedures. Images capturing the cervical region, known as *cervigrams*, are the gold-standard for the basic evaluation of cervical cancer presence. Cervigrams have high inter-rater variability especially among less skilled medical specialists. In this paper, we develop a fully-automated pipeline for cervix detection and cervical cancer classification from cervigram images. The proposed pipeline consists of two pre-trained deep learning models for the automatic cervix detection and cervical tumor classification. The first model detects the cervix region **1000** times faster than state-of-the-art data-driven models while achieving a detection accuracy of **0.68** in terms of intersection of union (IoU) measure. Self-extracted features are used by the second model to classify the cervix tumors. These features are learned using two lightweight models based on convolutional neural networks (CNN). The proposed deep learning classifier outperforms existing models in terms of classification accuracy and speed. Our classifier is characterized by an area under the curve (AUC) score of **0.82** while classifying each cervix region **20** times faster. Finally, the pipeline accuracy, speed and lightweight architecture make it very appropriate for mobile phone deployment. Such deployment is expected to drastically enhance the early detection of cervical cancer in less-developed countries.

1. Introduction

Cancer arises in the body when the cells of a specific organ start to grow abnormally. This abnormal cell growth can affect different body organs in women like the breast and the cervix. Cervical cancer is the type of cancer that affects the cervix of a woman. The cervix is the neck-shape passage at the bottom of the uterus. In developing countries, cervical cancer is ranked third as the most fatal type of cancer Torre, Siegel, Ward, and Jemal (2016). In 2012, almost half a million cases of cervical cancer were reported worldwide. Half of these cases were fatal Torre et al. (2016). Globally, more than 700 daily deaths due to cervical cancer are reported LaVigne, Triedman, Randall, Trimble, and Viswanathan (2017). These numbers seem to be only rising as it is expected that the number of fatalities will reach 400,000 annually by the year 2030 Wittet, Goltz, and Cody (2015). Cervical cancer screening is the process by which a test is performed to check the existence of abnormal tissues or cancerous cells in the cervix. Screening can help curing cervical cancer by detecting Cervical Intraepithelial Neoplasia (CIN) which indicates the abnormal change in the cervix. According to the world health organization (WHO), CIN can be cast into three types: CIN1 (mild), CIN2 (moderate) and CIN3 (severe). While, a CIN1 infection needs observation only, infections classified as CIN2 and CIN3 types require a specialized treatment. Physicians rely on different screening methods to differentiate between these types to decide if the patient needs treatment or not. Current screening methods include *PAP* test, attributed to George Nicholas Papanicolaou Tan and Tatsumura (2015), testing of a specific human Papillomavirus (HPV) and visual inspection Mayrand et al. (2007). *PAP* test defines the process of taking a sample from the cervix and inspecting it under the microscope Tan and Tatsumura (2015). However, the *PAP* test is impaired with false negative rates ranging from 6% to 55% Koss (1989). The HPV test is a DNA test that detects cervical cancer by associating it with a specific HPV type. Usually this test is not recommended as it suffers from a high false positive rate Hartman, Hall, Nanda, Boggess, and Zoulnoun (2002). Furthermore, the cost of such tests is quite high. In developing countries, it

* Corresponding author.
E-mail addresses: g201080740@kfupm.edu.sa (Z. Alyafeai), lahouari@kfupm.edu.sa (L. Ghouti).

https://doi.org/10.1016/j.eswa.2019.112951
0957-4174/

is difficult to afford these tests hence they rely on visual inspection. However, visual inspection can be tricky and requires expertise which lacks in such countries. Cervix shape, color and texture can help physicians decide which treatment to be taken thereafter Jordan (2009). Hence, detecting these types is related to the expertise of the physician which is not available in developing countries. *Digital cervicography* refers to the process of taking a photograph of the cervix called (cervigram) after applying 5% acetic acid. Hence, automated detection and classification can be used on cervigrams. Recently, deep learning models have shown impressive effectiveness in object detection and classification.

1.1. Paper outline

The rest of the paper is organized as follows. Section 2 gives an overview of existing solutions for the detection and classification of cervical cancer. Then, the proposed deep learning model is introduced in Section 3 where the detection and classification modules are discussed in details. Performance analysis of the proposed pipeline is presented in Section 4. The performance of the RoI detection and cancer classification modules is compared to state-of-the-art models in the literature. Computational efficiency, detection and classification accuracy are contrasted to highlight the superiority of the proposed deep learning pipeline. Section 5 concludes the paper where future work directions are given and conclusions drawn.

2. Literature review

2.1. Existing techniques using hand-crafted features

In Mango (1994), Mango discussed the use of computer-based algorithms to detect cancerous cells in the cervix. He recommended the *PAPNET* cytological screening system for the detection of abnormal cells Cresce and Lifshitz (1991). Mango's solution is based on a conventional PAP smear test and an artificial neural network (ANN) model. The ANN model allows the automatic detection of the precancerous tests. More elaborate solutions require a detection module to extract the cell nuclei region from cervigrams. For instance, Bamford and Lovell employed an active contour method to extract the cell nuclei region Bamford and Lovell (1998). Bamford and Lovell identified the RoI are prior to extracting a specific number of contours. Then, only the most relevant contours are retained for further processing Bamford and Lovell (1998). Feature extraction schemes have been also recommended in the literature. A conventional region growing algorithm is modified by Mat-Isa et al. Mat-Isa, Mashor, and Othman (2005). The proposed modification, called the seeded region growing features extraction (SRGFE), infers the size and grayscale levels of a specific RoI in the cervigram image Mat-Isa et al. (2005). Another approach is attributed to Chang et al. where the size and deformation of a cell nuclei are used to categorize it as abnormal Chang et al. (2009). Chang et al. pre-processed the cervigram image to remove the noisy parts prior to the extraction of the cell nuclei region Chang et al. (2009). Then, two complementary approaches are suggested to classify the cells using the grayscale level and energy, respectively. The resulting classifier is able to discriminate the abnormal cells. A data-driven solution is described in Kim and Huang (2013) where Kim and Huang used an optimized bounding box (OBB) method to detect the RoI in cervigram images. The RoI is usually a BB rectangle centered around the cervix. Several BB regions with different locations and scales are extracted from similar images using a similarity metric. Then, the "*best*" BB region is retained using a combination of Euclidean distance and intersection over union (IoU) metrics. Finally, a two-variant classifier is built on majority voting method and support vector machine (SVM) model. Both classifiers are trained using engineered (or hand-crafted) features. These features are constructed using cervical colour and texture representations. Song et al. used first a Sobel filter to detect the cervix RoI Song et al. (2015a). Then, a multi-modal approach is proposed for the cervical cancer classification. The latter approach collects information from the cervigrams and the clinical tests. This classification decision is based on the collected information. Song et al. evaluated a similarity measure based on the classifier information to extract a label for the cervigram under consideration Song et al. (2015a). It is noteworthy to mention that Kim and Huang assessed the performance of two different categories of cancer classifiers Kim and Huang (2013). In the first category, classifiers are designed and trained using hand-crafted features. These features, well tested in the computer vision field, are standard image descriptors. Such descriptors include pyramids of local binary patterns (PLBP), pyramid color histogram in the L*a*b* color space (PLAB) and pyramid histogram of orientated gradients (PHOG) Davies (2012). In the second classifier category, auto-extracted features are used to train the cervical cancer classifier. These features are directly inferred during learning by a random forest, SVM or convolutional neural network (CNN) model.

Using a publicly available patient data, Adem et al. classified four different target variables including the Schiller, Citology, Biopsy and Hinselmann features Adem, Kiliçarslan, and Cömert (2019). These features represent potential cervical cancer risks. Unlike our proposed model, Adem et al. used a small sized dataset consisting of patient historical data. The deep learning classifier consists of a stacked autoencoder model. Correct classification rates reached 0.978 which highlights the superiority of deep learning models in patient diagnostic support systems. In Fernandes, Cardoso, and Fernandes (2017), Fernandes et al. predicted cervical cancer risks and subjective quality assessment scores from colposcopic images based on different modalities and human experts. The prediction model relies on a regularization-based transfer learning strategy Goodfellow, Bengio, Courville, and Bengio (2016). Such strategy enables the source and target models to share the same model parameters and reduces the size of the required training data. In addition, the model trained on one expert/modality subset can be easily extended thanks to the transfer learning strategy. Using only nucleus-level texture features, Phoulady et al. Phoulady, Zhou, Goldgof, Hall, and Mouton (2016) classified cervical tissue as normal or cancer. The proposed features are processed using a two-step nucleus-level analysis. In the first step, nucleus-level information is captured using an adaptive multilevel thresholding segmentation. Then, the shape of the segmented regions is approximated using an ellipse fitting algorithm. The two-step classifier, called adaptive nucleus shape modeling (ANSM) algorithm, achieved a classification accuracy of 0.933 with zero false negative rates. Thanks to its performance, the ANSM model can improve the overall performance of cervical histology image analysis tools.

Devi et al. surveyed and analyzed several ANN architectures used in the classification of cervical cancer Devi, Ravi, Vaishnavi, and Punitha (2016). The most efficient architectures are recommended for the binary classification of cervical images as normal or abnormal cervical cells. The performance of these architectures are contrasted to that of the manual screening methods such as the PAP smear and liquid cytology based (LCB) tests.

To mitigate the image segmentation challenges, Zhang et al. proposed a segmentation-free classifier for cervical cancer Zhang et al. (2017). Unlike previous models, Zhang et al. trained a deep learning model using automatic features. To cope with the limited size of the training data, Zhang et al. trained their model using a two-stage approach. In the first stage, the deep learning model is pre-trained on a natural image dataset. Then, the trained model is tuned using nuclei-centered patches extracted from

Table 1
Summary of existing cervical cancer detection and classification techniques.

Authors	Dataset Size	Detection	Classification	Sensitivity (%)	Specificity (%)
Xu et al. (2017b)	1112	OBB method	Feature-based CNN	80.87 ± 7.43	75.94 ± 7.46
Song et al. (2015a)	280	Data-driven and BB similarity	Multimodal with clustering	83.21	94.79
Kim and Huang (2013)	2000	OBB method	Majority vote	73.00	77.00
Sankaranarayanan et al. (2004)	54, 981	Not available	Screening	79.00	86.00
Denny, Kuhn, Pollack, and Wright (2002)	2754	Visual Inspection	Visual Inspection	70.00	79.00

adaptively resampled cervical images. Despite its good performance, this model requires accurately extracted image patches.

Another deep learning model for cervical cancer classification is attributed to Almubarak et al. Almubarak et al. (2017). In this deep learning classifier, 4-class grades of CIN images are considered where the epithelium region is partitioned into 10 segments. Then, each of these segments is further split into 3 blocks where each block is classified using a CNN model. Finally, the obtained scores are fused to classify the segments and the whole epithelium. Almubarak et al. assessed their model using a small dataset consisting of 65 cervical images only where they achieved 0.7725 accuracy.

A new cervical image dataset is released by Xu et al. for benchmarking purposes Xu et al. (2017a). Unlike previous datasets, expert annotations and diagnoses are also made available. In this way, image-based cervical disease classification algorithms can be fairly evaluated and compared. In addition, hand-crafted features extracted from the dataset images are also provided. These features are similar to the ones discussed in Appendix A.

The work proposed in Almubarak et al. (2017) is extended by AlMubarak et al. using a hybrid deep learning model AlMubarak et al. (2019). Hybrid features extracted from 10 vertical blocks in the epithelium region. Feature extraction is performed using 27 hand-crafted and automatic features. The latter features are obtained using a sliding window-based CNN model associated with each vertical block. The hybrid model achieved 0.807 accuracy in the classification of the cancer grade of the whole epithelium region.

Bhargava et al. suggested the use of hand-crafted features to classify cervical cancer Bhargava, Gairola, Vyas, and Bhan (2018). In their classification solution, Bhargava et al. trained 3 different classifiers using the proposed hand-crafted features including SVM, KNN and ANN models. These classifiers are trained using 66 cervical images only. The classification performance varied considerably among these classifiers where accuracy scores ranged from 0.621 to 0.955.

Another review of current cervical cancer pattern classification techniques is attributed to Kudva et al. Kudva and Prasad (2018). The reviewed techniques focused on using image pattern classification solutions to classify cervical images using features related to color, vascular pattern and lesion margin information. It is well known that these features have good discriminative power with respect to normal and abnormal lesions Song et al. (2015b).

Table 1 summarizes the main techniques proposed for the detection and classification of cervical cancer. Unlike our proposed solutions for detection and classification, data-driven techniques such as those proposed by Xu et al. Xu et al. (2017b), Song et al. Song et al. (2015a) and Kim et al. Kim and Huang (2013), suffer from high computational complexity as they require matching against all cervigram images available in the dataset. Such techniques may be considered as non-parametric Friedman, Hastie, and Tibshirani (2009). Zhang et al. Zhang et al. (2017) achieved high accuracy and area under the curve metrics using a fine-tuned pretrained ConvNet but the results are evaluated on cytology images which contain simpler features compared to cervigrams. Hu et al. Hu et al. (2019) used a combination of Faster R-CNN for detection and a pretrained CNN model for classification; however region based methods are considered slow compared to single shot detectors like (you only look once) YOLO. A hybrid deep learning approach is suggested using deep learning approaches and handcrafted approaches on cervical cancer digital histology by AlMubarak et al. AlMubarak et al. (2019). Kudva et al. used a shallow CNN for the classification of cervical cancer however their appraoch is not fully automated as they manually extracted patches of size 15×15 from the cervigram images Kudva, Prasad, and Guruvare (2018). Sornapudi et al. applied deep learning for the detection of nuclei in histology images using super pixels clustering and training with convolutional neural network (CNN) Sornapudi et al. (2018).

2.2. Image segmentation and object detection techniques

Image segmentation techniques attempt to divide an input image into different regions according to color uniformity, texture richness and edge activity. These Techniques are further categorized into segmentation approaches based on: 1) image thresholding schemes like Otsu algorithm Kapur, Sahoo, and Wong (1985); 2) region growing techniques Adams and Bischof (1994)-Tang (2010); 3) image clustering Coleman and Andrews (1979)-Chuang, Tzeng, Chen, Wu, and Chen (2006) and 4) tree partitioning schemes Salembier and Garrido (2000). Otsu algorithm picks the threshold that minimizes the *intra-class* variance Kapur et al. (1985). The seeded region growing algorithm is attributed to Adams and Bischof Adams and Bischof (1994). Then, this algorithm was extended to color image segmentation by Tang Tang (2010). While similar color channels are grouped together by Coleman and Andrews in Coleman and Andrews (1979), images are segmented using a fuzzy c-means clustering solution in Chuang et al. (2006). The use of tree-partitioning in image segmentation was first suggested by Salembier and Garrido Salembier and Garrido (2000). A binary partition tree is used to efficiently represent and segment images Salembier and Garrido (2000). Felzenszwalb et al. proposed the use of efficient graph-based clustering solutions to segment images and detect objects Felzenszwalb and Huttenlocher (2004). It should be mentioned that most of deep learning object detection algorithms use the *selective search* algorithm as a starting point in their detection process van de Sande, Uijlings, Gevers, and Smeulders (2011)-Uijlings, van de Sande, Gevers, and Smeulders (2013). To overcome the limitations of *exhaustive search*, Uijlings et al. proposed a selective search strategy along with rich features and computationally-expensive classifiers to detect objects in images van de Sande et al. (2011)-Uijlings et al. (2013). Uijlings et al. algorithm is usually initialized using the Felzenszwalb et al. scheme Salembier and Garrido (2000). Uijlings et al. treated image segmentation as a selective search problem where approximate image locations are selected over a limited number of objects localized inside bounding boxes Uijlings et al. (2013). The first deep learning object detection algorithm, commonly known as region-CNN (R-CNN), uses rich feature hierarchies to accurately detect an object and segment an image semantically Girshick, Donahue, Darrell, and Malik (2014). The R-CNN model consists of 3 different modules. *Region propos-*

als using the *selective search* algorithm van de Sande et al. (2011)-Uijlings et al. (2013) is carried out by the first module. The second module, a CNN model, generates self-extracted features for each proposed region. Finally, the third module performs the classification task using an SVM classifier model. To speed-up the detection process, an image RoI is first extracted from the input image as proposed by Girshick Girshick (2015). The fast R-CNN is enhanced further by Ren et al. in Ren, He, Girshick, and Sun (2015). Ren et al. algorithm, known as faster R-CNN, merges the region proposal and CNN modules to reduce the overall computational complexity. Real-time object detection has been made possible thanks to the *you only look once* (YOLO) algorithm Redmon, Divvala, Girshick, and Farhadi (2016); Redmon and Farhadi (2017). Unlike previous detection algorithms, *YOLO* algorithm and its improved versions (*YOLO 9000* and *YOLO v3* formulate the object detection problem as a regression one. Then, the regression problem is solved to assign detection probabilities to distinct objects located separately. In addition, the input image is scanned only once through a single ANN model. An instance-based method, called the *mask R-CNN*, extends the faster R-CNN model by merging the object prediction and bounding box detection modules He, Gkioxari, Dollár, and Girshick (2017). Real-time detection operates at 45 frames per second in the baseline *YOLO* model and 155 frames per second in the improved version Redmon et al. (2016); Redmon and Farhadi (2017). Finally, the *YOLO* algorithm can be easily generalized to detect objects in images found in various applications such as medical and space imaging.

2.3. Image classification techniques

Image classification has been the focus of active research in the computer vision and image processing fields Gonzalez and Woods (2018); Szeliski (2011). The pioneering experiments of Hubel and Weisel on the visual cortex of cats have paved the way for the design of state-of-the-art computer vision algorithms Hubel and Wiesel (1963). In their experiments, Hubel and Weisel realized that some neurons in the cat's brain are stimulated by edges regardless of the position Hubel and Wiesel (1963). Since then, computer vision researchers have been continuously proposing models and algorithms that focus on image edges given their importance to the human visual system (HVS). Fukushima proposed the first computer system, called the *neocognitron*, to model the visual cortex. The *neocognitron* model is a layered structure with local receptive fields to activate a specific region at a time Fukushima and Miyake (1982). Then, Rumelhart et al. trained an ANN model by back-propagating (BP) the errors Rumelhart, Hinton, and Williams (1986). In his excellent review blog, Schmidhuber gives a detailed historical accounts of the major developments of the back-propagation algorithm Schmidhuber (2018). Following the major successes of the BP algorithm in training ANN models, Lecun et al. used the BP algorithm to train a deep structure of convolutional layers called *LeNet-5* LeCun, Bottou, Bengio, and Haffner (1998). In 2012, a major breakthrough of deep learning models in the computer vision field was accomplished by Krizhevsky et al. Krizhevsky, Sutskever, and Hinton (2012). *AlexNet*, a deep CNN model proposed by Krizhevsky et al., achieved the lowest top 5 error score of 15.4% in the 2012 *ImageNet* image classification challenge Krizhevsky et al. (2012). In this challenge, classifier models are tested with 1 million images pertaining to 1000 categories. The top 5 error specifies that the test image is correctly classified if it is among the 5 highly scored classes. To outperform their competitors, Krizhevsky et al. used 5 convolutional layers followed by 2 fully-connected in their *AlexNet* model. In addition, pooling, normalization and non-linear activation layers are also used Krizhevsky et al. (2012). Following its success, Zeiler and Fergus introduced some modifications into the *AlexNet* model to win the 2013 *ImageNet* challenge by a top 5 error score of 14.8% Zeiler and Fergus (2014). The *ZFNet* model, proposed by Zeiler and Fergus, used bigger filters to carry out the convolution operations Zeiler and Fergus (2014). In 2014, a different approach was adopted by the Visual Geometry Group (VGG) at Oxford University to win the *ImageNet* challenge Simonyan and Zisserman (2014). In fact, Simonyan and Zisserman, used a very deed CNN model with 3×3 convolution filters only. Simonyan and Zisserman model, called *VGGNet*, attained the lowest top 5 error score of 9.33% in the 2014 *ImageNet* challenge Simonyan and Zisserman (2014). In the same year, a research team from Google developed the *GoogLeNet* model for the *ImageNet* challenge Szegedy et al. (2015). Szegedy et al. introduce the inception layers in the *GoogLeNet* model using less parameters. *GoogLeNet* outperformed the *VGGNet* and scored a top 5 error of 9.13%. The newly-introduced inception layers contained convolution filters with varying sizes and max pooling layers with a concatenation filter at the end Szegedy et al. (2015). Another model, called *ResNet*, was introduced by the research team at Microsoft in 2015 He, Zhang, Ren, and Sun (2016). The *ResNet* model contained up to 152 layers and *skip connections*. At this depth, *ResNet* is 8 times deeper than the *VGGNet* model. Skip connections bypass some convolutional layers to mitigate the effect of vanishing gradients as less multiplications will be involvedHe et al. (2016). At that year, the *ResNet* model achieved the lowest top 5 error scores between 6.7 and 5.7 in the *ImageNet* competition He et al. (2016). In the 2016 version of *ImageNet* challenge, an ensemble model, consisting of 5 pre-trained CNNs, attained the lowest top 5 error of 2.99%. The pre-trained models consisted of *Inception-v3, Inception-v4, Inception-ResNet-v2, Pre-Activation ResNet-200*, and *Wide ResNet (WRN-68â2). ResNeXt*, attributed to a Facebook research team, achieved a top-5 error of 5.3% in the 2016 *ImageNet* challenge thanks to residual layers Xie, Girshick, Dollár, Tu, and He (2017). The *ResNeXt* is a highly modular network architecture where a single block is duplicated several times. Thanks to this duplication, *ResNeXt* requires less hyper-parameters to tune. To win the 2017 version of the *ImageNet* challenge, Hu et al. proposed a deep learning model called the *squeeze-and-excitation networks* (SEN) Hu, Shen, Albanie, Sun, and Wu (2017). The SEN model successfully reduced the top-5 error to 2.251% which represents a 25% relative improvement over the winning model of the previous year challenge. In the SEN model, a *squeeze-and-Excitation* (SE) block explicitly models the inter-dependencies between the image channels to adaptively recalibrate the channel-wise feature responses. A historical summary of the deep learning achievements in the *ImageNet* image classification challenge is outlined in Fig. 1. The 2010 and 2011

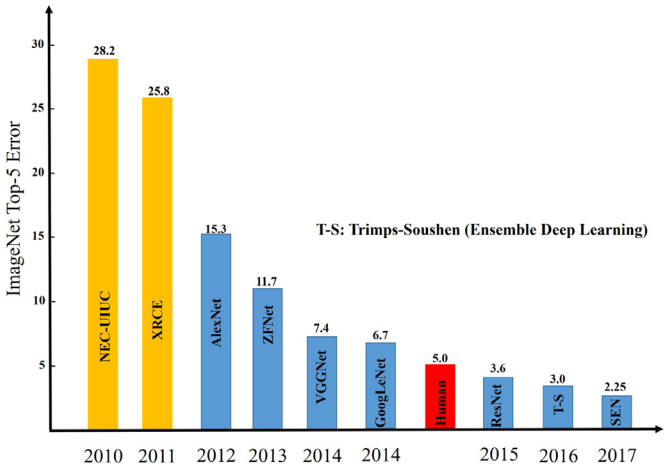

Fig. 1. Main breakthroughs in *ImageNet* image classification challenge.

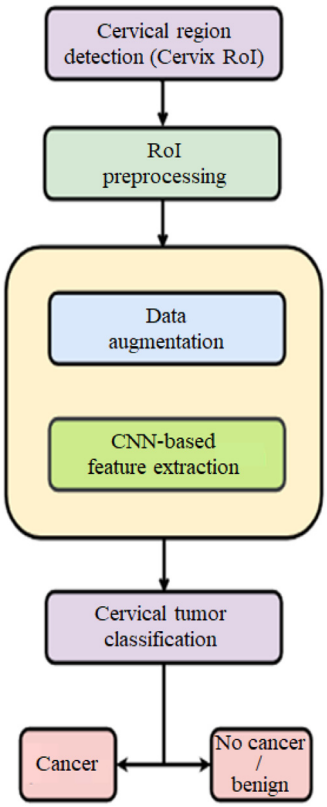

Fig. 2. Architecture of the proposed deep learning pipeline.

winning solutions, shown in Fig. 1, are not based on deep learning solutions. The 2012 challenge version witnessed the dominance of the deep learning models in the *ImageNet* image classification challenge. Moreover, the *ResNet* model did not only outperform its competitors in 2015 but it surpassed the human performance in classifying still images. This landmark breakthrough has paved the way for next generation image classifiers Hu et al. (2017).

3. Proposed fully-automated deep learning pipeline

To alleviate the need for human medical expertise, we propose a fully-automated deep learning pipeline for the detection of cervix regions and classification of cervical tumors. To the best of our knowledge, this is the first pipeline of its kind where human intervention is not required to localize the cervical RoI block or diagnose the presence of cervical cancer. Fig. 2 provides a schematic outline of the architecture of the proposed deep learning pipeline.

The crucial components of the proposed pipeline are:

1. Cervix detection and cervical RoI extraction.
2. RoI pre-processing and data augmentation.
3. Automatic feature extraction.
4. Classification of cervical tumors.

The components of the proposed pipeline are thoroughly discussed below.

3.1. Cervix detection module

To extract the cervical RoI, a deep learning object detection model is proposed. State-of-the-art object detection can be achieved using *R-CNN* Girshick et al. (2014), *Fast-RCNN* Girshick (2015), *Faster-RCNN* Ren et al. (2015) or *YOLO* Redmon et al. (2016) algorithms. In our paper, we propose to design the RoI detection module using the *YOLO* algorithm as it performs real-time object detection at a speed of approximately 45 frames per second Redmon et al. (2016). Our cervix detection module is a modified version of the *GoogLeNet* image classifier model. As illustrated in Fig. 3, it consists of 24 convolutional layers followed by two fully-connected ones.

To reduce the number of convolutional layers without impairing the performance, the *inception* concept, attributed to Szegedy et al. Szegedy, Vanhoucke, Ioffe, Shlens, and Wojna (2016), is used in our cervix detection module. Inception layers, sub-modules attached to larger deep learning models, allow faster predictions while drastically reducing the total number of model parameters. Convolutional kernels are also factorized into smaller ones Szegedy et al. (2016). On the other hand, to ensure proper training without the risk of overfitting, skip or residual nodes are also used following a solution similar to that proposed in Szegedy, Ioffe, Vanhoucke, and Alemi (2017). In summary, our proposed detection module has a structure similar to that used in the *ResNet* model Szegedy et al. (2016).

The cervix detection module processes a cervigram image at a time. Features, extracted from the entire cervigram, are used to predict each bounding box around the cervical RoI area. The cervigram image is first divided into $S \times S$ grid with equal area of each grid. Then, every grid predicts B bounding boxes and their corresponding confidence scores. These scores quantify the model confidence about the existence of a the cervical RoI area inside the bounding box. The confidence score, $conf\_score$, is defined as:

$$conf\_score = Pr(Object) \times IoU_p^t \quad (1)$$

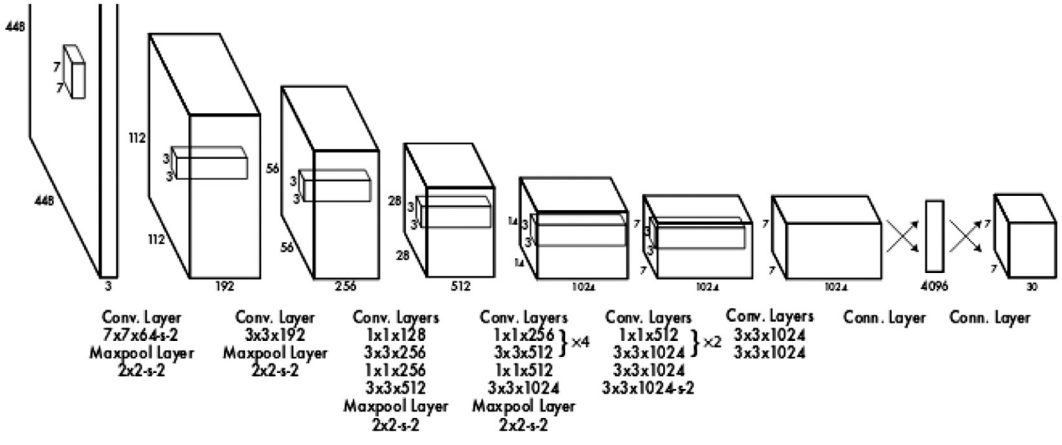

Fig. 3. Architecture of cervix detection module.

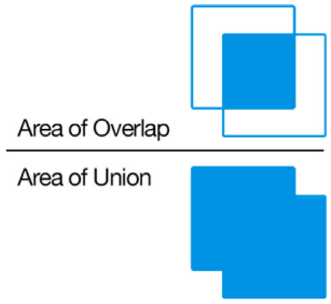

Fig. 4. IoU computation in detection module.

where IoU_p^t specifies the intersection over the union (IoU) of the true and predicted bounding boxes, b_t and b_p, around the cervical RoI area. Given b_t and b_p, their IoU is defined as follows:

$$IoU_p^t = \frac{b_t \cap b_p}{b_t \cup b_p} \tag{2}$$

Eq. (2) evaluates the overlapped area of the true and predicted bounding boxes with respect to their union. The IoU_p^t score takes values in the interval [0, 1]. An IoU_p^t score of 1 would mean perfect object detection and a totally missed object detection would result in an IoU_p^t score of 0. The IoU metric computation is demonstrated in Fig. 4.

Each bounding box is associated with 5 parameters to represent its spatial location and the object confidence score $conf\_score$. The spatial location of the bounding box includes the (x, y)-coordinates of its upper left corner, its width, w, and height h. Each grid cell evaluates C conditional probabilities $Pr(Class_i|Object)$. $Pr(Class_i|Object)$ gives the confidence of the model in assigning the detected object to the ith object class. Then, the class specific probabilities are evaluated using:

$$\begin{aligned} &Pr(Class_i|Object) \times Pr(Object) \times IoU_p^t \\ &= Pr(Class_i) \times IoU_p^t \end{aligned} \tag{3}$$

Finally, the bounding box predictions are encoded as $S \times S \times (5 \cdot B + C)$ as each of the B bounding boxes is represented using 5 parameters assuming C different object classes.

The loss function, adopted from the *YOLO* algorithm, is defined as Redmon et al. (2016):

$$\begin{aligned} Loss = {} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} \left(C_i - \hat{C}_i\right)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \mathbb{1}_{i}^{obj} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2 \end{aligned} \tag{4}$$

where $\mathbb{1}_i^{obj}$ and $\mathbb{1}_{ij}^{obj}$ are indicator functions to denote the presence of the object obj in the ith grid cell and the prediction responsibility of the jth bounding box in the ith grid cell, respectively. True and predicted centers of the bounding box are given by the (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ pairs. The width and height of true and predicted bounding boxes are defined by w_i, h_i, $\hat{w}_i$ and $\hat{h}_i$, respectively. λ_{coord} and λ_{noobj} are weight parameters to scale up and down the losses due to object localization and classification. True and predicted object classes are given by C_i and $\hat{C}_i$, respectively. $p_i(c)$ and $\hat{p}_i(c)$ represent the true and predicted probabilities of the ith class object being localized in the bounding box. In summary, Eq. (4) evaluates the prediction error between the true and predicted bounding boxes. This cost minimization steps are summarized in Fig. 5.

3.2. RoI pre-processing and data augmentation

3.2.1. RoI Pre-Processing

The detected cervical RoI areas are resized to a standard size of $256 \times 256 \times 3$. Each of the resized images is zero-centered and normalized using the z-scaling approach Li, Karpathy, and Johnson (2016):

$$RoI_{zs} = \frac{RoI - \mu_{RoI}}{\sigma_{RoI}} \tag{5}$$

where RoI_{zs}, μ_{RoI} and σ_{RoI} represent the z-scaled RoI image, the mean and standard deviation of RoI images. Eq. (5) is applied to all RoI images where each color channel is processed separately.

3.2.2. Data augmentation

Since the number of available cervigram images is very limited, training of any deep learning model using these images will not be carried out properly. In fact, deep learning models perform poorly when trained using moderate training data sizes as shown in Fig. 6. A low data sizes, traditional machine learning algorithms can even outperform their deep learning counterparts as evidenced by Fig. 6.

Therefore, to enable our deep learning pipeline to operate in the high training size regime, cervigram images will be generated artificially using the *data augmentation* approach. Data augmentation applies several mild image transformations on the training cervigram RoI images in order to increase the learning ability of our deep learning models Chatfield, Simonyan, Vedaldi, and Zisserman (2014). For instance, to train the *AlexNet* model, Krizhevsky et al. used image translations and horizontal reflections on the available training images Krizhevsky et al. (2012). In this way, they were able to increase the number of training image by 2048. In our case, we will use different mild image alterations including like random cropping, random flipping and random rotation. Sample augmented RoI cervigram images are illustrated in Fig. 7.

Finally, the normalized and augmented RoI images are fed into the subsequent modules for automatic feature extraction and cervical tumor classification.

3.3. Automatic feature extraction module

A lightweight convolutional model is proposed to automatically extract features from the cervigram RoI images. Color-based tensors of sizes $256 \times 256 \times 3$ are fed to the feature extraction module. Unlike the models based on hand-crafted features, the standard RGB color space is used in our module. To allow adequate training, N training RoI samples are processed at each step using the batch normalization approach Ioffe and Szegedy (2015).

Our proposed feature extraction module has two distinct architectures:

1. **Model 1:** Two convolutional layers are used along with 3×3 convolutional kernels and 2×2 strides. 16 different kernels are used in the first convolutional layer and 32 in the second one. Each convolutional layer is followed by a ReLU activation and max-pooling layers. Feature vector sizes drop by half after the last two layers. Then, the resulting feature maps are flattened into a dense layer of 128 neurons and processed through a hyperbolic tangent activation layer.
2. **Model 2:** Three convolutional layers are used along with 3×3 convolutional kernels and 2×2 strides. The number of kernels in these layers is 16, 32 and 64, respectively. The remaining layers have the same structure as **Model 1** above.

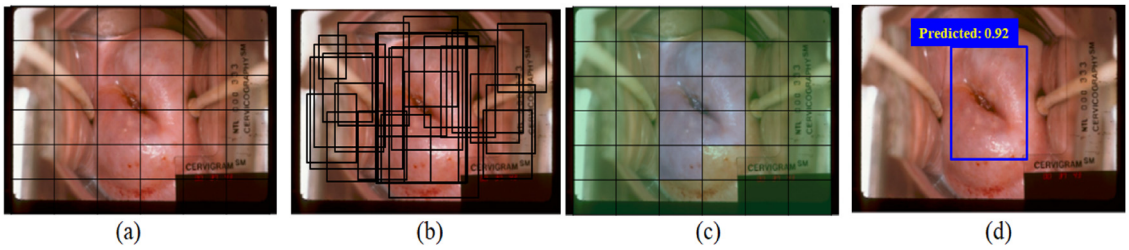

Fig. 5. Cerviccal RoI detection process. Equal-sized grid cells (a). Bounding boxes and confidence scores (b). Class probability map (c). Detected RoI area (d).

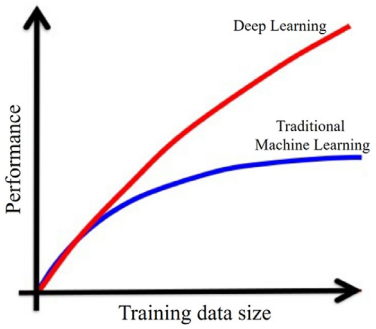

Fig. 6. Effect of training data size on deep learning performance.

Fig. 8 illustrates the proposed feature extraction process. The lightweight configuration of the feature extraction module makes it very suitable for training instances with low and moderate number of training samples. Moreover, the cervical cancer problem does not only suffer from the lack of adequate training samples but also data imbalance.

Sample automatic features, extracted from the training RoI images, are shown in Fig. 9.

3.4. Cervical cancer classification module

Once the automatic features are extracted, they are *flattened* into 128-dimensional feature vectors. The flattened feature vectors are processed through a logistic layer to produce a probability score about the presence or absence of cervical cancer. In the case of multiple decisions, a *softmax* layer is used instead to assign class probability scores Goodfellow et al. (2016). The sigmoid layer maps the feature map vector to a real-valued scalar between 0 and 1. Then, the loss function, a cross-entropy measure, is estimated as follows Goodfellow et al. (2016):

$$cross\_ent = -y \log \cdot (p) - (1 - y) \cdot \log (1 - p) \quad (6)$$

where y is the true class label (0 or 1) and p is the output of the logistic layer. p represents a probability score assigned to the likelihood of the presence of cervical cancer. The loss function of Eq. (6) is minimized in an iterative fashion where its gradient is back-propagated to update the parameters of the classifier module shown in Fig. 10.

Unlike existing models based on hand-crafted features, the proposed classifier module is seamlessly integrated with the automatic feature extraction module. This integration allows simple deployment of the overall pipeline and considerable boost in the computational efficiency. Therefore, the hidden layers, shown in Fig. 10, can be replaced with the 2 or 3 convolutional layers of the feature extraction module.

4. Performance evaluation of proposed deep learning pipeline

4.1. Experiment design and setup

All computer experiments are carried out on a computer workstation running 16.04 Linux Ubuntu operating system. An Invidia graphical processing unit (GPU), the GeForce GT 740M with 2 GB memory, is used to speed up the deep learninng computations. All deep learning models are implemented in the *Keras* deep learning package with Google *TensorFlow* backend using Python 3 programming language. In addition, the Google *Colaboratory* deep learning web service was used as it offers 8 GB of free GPU memory for research purposes.

4.2. Cervigram datasets

Two different cervigram datasets are used to train and evaluate the main modules of the proposed deep learning pipelines.

4.2.1. Intel&MobileODT dataset

The cervical cancer screening dataset provided jointly by Intel and MobileODT companies. The screening dataset is made publicly available in a machine learning international competition hosted by Google *Kaggle* portal Kaggle (2017a). The Intel&MobileODT dataset is primarily intended to train machine and deep learning models to correctly classify the cervix types. All training cervigram images are collected from women with no cervical cancer infection. These images are only used to train our RoI detection module. As these images are not intended for the training of object

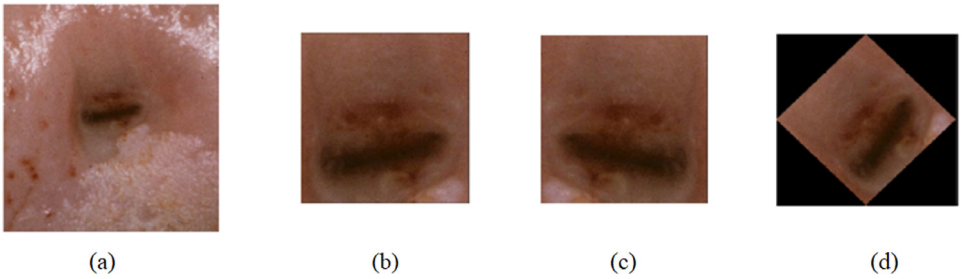

Fig. 7. Sample augmented RoI cervigram images. Original RoI (a). Randomly- cropped RoI (b). Horizontally-flipped RoI (c). Rotated RoI with 45^o. (d).

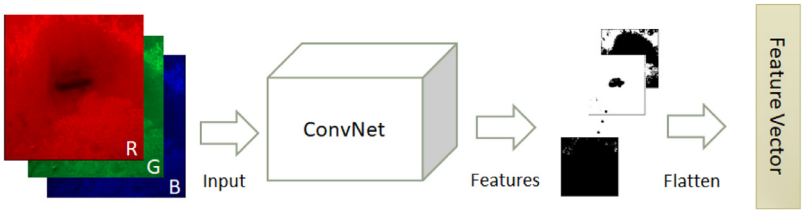

Fig. 8. Proposed feature extraction module.

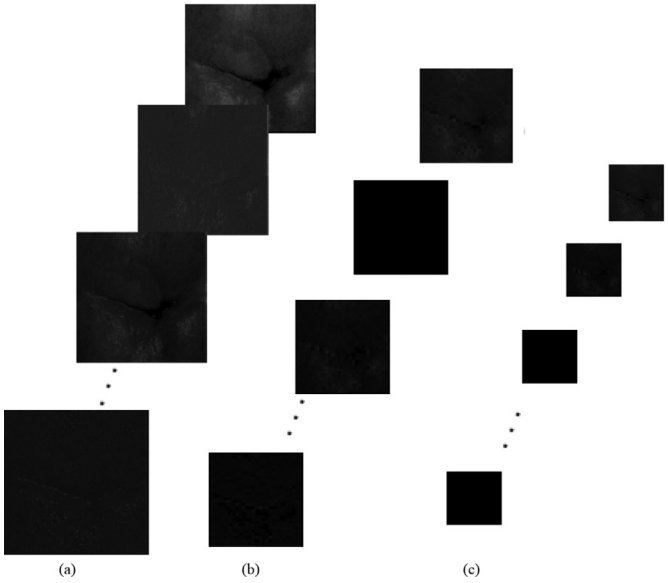

Fig. 9. Auto-extracted features from first (a), second (b) and third (c) convolutional layers.

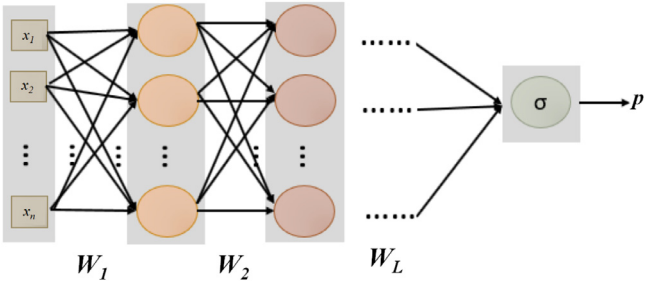

Fig. 10. Architecture of proposed cervical cancer classification module.

Table 2
Cervix types in Intel&MobileODT cervigram dataset.

Cervigram Type	Number of Images
Type 1	1241
Type 2	4348
Type 3	2426

detection deep learning models, the cervical RoI region must be manually annotated first. A human expert has manually annotated a considerable subset of the images in the dataset Kaggle (2017b). Sample annotated cervigrams are shown in Fig. 11.

The Cervix type is tightly related to the location of the transformation zone Jordan, Singer, Jones, and Shafi (2009). Table 2 gives a summary on cervix type distribution across the Intel&MobileODT dataset. It should be noted that only a subset of 1500 images are correctly annotated in this dataset Kaggle (2017b).

Fig. 11. Sample annotated cervigram images Kaggle (2017a,b).

Table 3
Cervigram dataset distribution for the cervix detection module.

Phase	Number of images
Training	1200
Test	300

Table 4
Available worst histology scores.

Label	Type
-2	No histolgoy
0	Normal
1	*CIN 1*
2	*CIN 2*
3	*CIN 3*
4	*CIN 4* (cancer)

For training, validation and test purposes, the annotated subset of the Intel&MobileODT dataset Kaggle (2017b) is split using a 80% - 20% ratio as indicated in Table 3.

4.2.2. NCI Guanacaste project dataset

The *Guanacaste* project cervical cancer dataset is provided by the American National Cancer Institute (NCI) Herrero et al. (1997). This dataset is used to train the feature extraction and cancer classification modules in our proposed deep learning pipelines. Throughout this project life cycle, cervigram images and patient medical history were collected from several locations in South America. The data, collected from 7000 patient visits, resulted in 44, 000 cervigram images. Available dataset information includes:

1. Patient age.
2. Worst histology.
3. Human papillomavirus (HPV) status.
4. Cervigram image.
5. Visit intervals in days.

During the histology test collection, only the worst histology diagnosis is retained for each patient. Table 4 summarizes the worst histology test scores available in the *Guanacaste* project dataset. A 0 classification label is assigned to normal cervigrams. Cases with a mild intraepithelial neoplasia (coded as *CIN 1*) are given a class

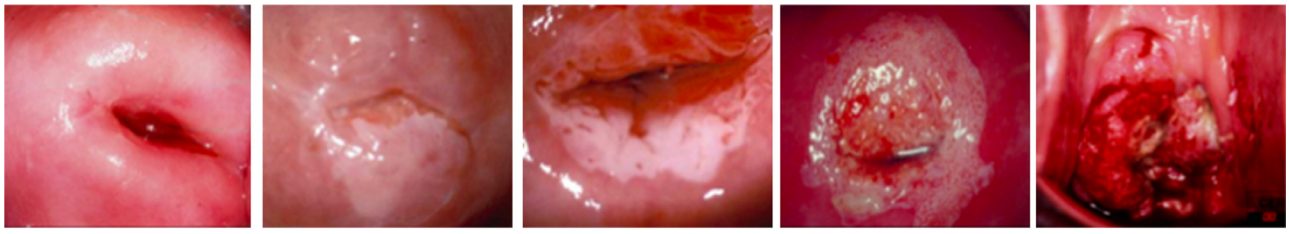

Fig. 12. Sample histology cases Herrero et al. (1997). Normal (a). *CIN 1* (b). *CIN 2* (c). *CIN 3* (d). *CIN 4* (cancer) (e).

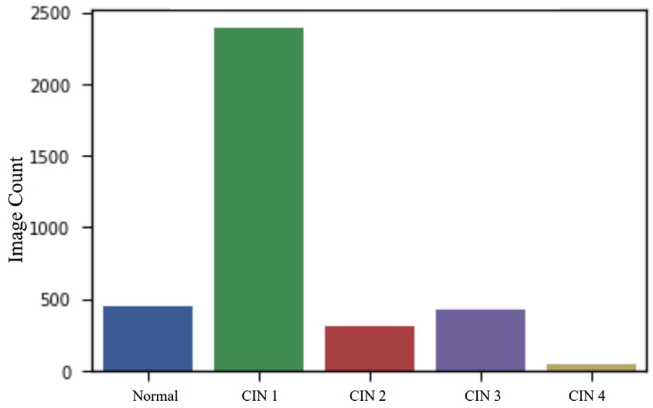

Fig. 13. Class distribution across all labeled images in the *Guanacaste* project dataset.

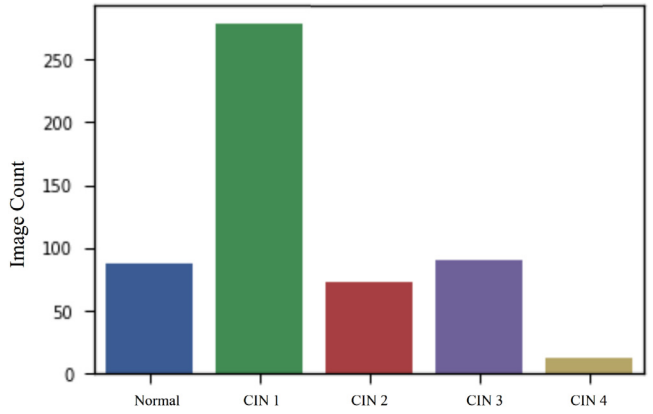

Fig. 14. Class distribution in filtered images in the *Guanacaste* project dataset.

label of 1. In general, *CIN 1* cases do not require any treatment. However, cases with moderate and severe intraepithelial neoplasia (*CIN 2* and *CIN 3*) are assigned to class labels 2 and 3, respectively. Further treatment is usually recommended to patients diagnosed as *CIN 2* and *CIN 3* cases. Finally, cervical cancer cases, encoded as *CIN 4*, are assigned to class label 4. Sample cases, with various cancer grades, are provided in Fig. 12.

It should be noted that the majority of the cervigram images do not have a valid wort histology record. The data distribution of the labeled cervigrams is shown is Fig. 13. In addition, many subjects are represented with multiple cervigram images in the dataset. It is clear from Fig. 13 that this data is highly imbalanced towards the *CIN 1* class. To avoid overfitting, we kept only the cervigram with worst histology per patient. The resulting class distribution is depicted in Fig. 14.

Although there is a substantial decrease in the number of images per cancer grade class, data imbalance is still impairing the

Table 5
Balanced cervigram dataset distribution among cancer classes.

Cervigrams / Class	Normal-*CIN 1*	*CIN 2-CIN 4*
Total number	174	174
Training	157	157
Test	17	17

Table 6
Balanced and augmented cervigram dataset distribution among cancer classes.

Phase / Class	Class 0	Class 1
Training	628	628
Test	17	17

filtered dataset. To remedy this impairment, we will select only a representative balanced data subset where all classes are equally represented. In this balanced subset, all cervigram images are scaled to an image size of 2891×1973 pixels. Table 5 gives a summary of the class distribution among the retained images in the data subset. Data split into training and test subsets using a 90% to 10% ratio would lead to 157 training and 17 test images, respectively. It is clear that 157 images will not allow adequate training of our cervical cancer classifier module. Hence, we will need to rely on the data augmentation technique introduced in Section 3.2.2.

For classification purposes, the cancer grade classes are cast into two distinct classes as follows:

1. **Class 0**: As normal and *CIN 1* types are very unlikely to be cancerous, they are grouped together in the cancer-negative class.
2. **Class 1**: Given their cancer risk likelihood, *CIN 2-CIN 4* types are assigned to the cancer-positive class.

Cervigram samples pertaining to the negative and positive cancer classes are presented in Fig. 15.

To appreciate the challenges in classifying the *Guanacaste* project dataset, the t-distributed stochastic neighbor embedding (*t-SNE*) algorithm is used for visualization purposes Maaten and Hinton (2008). The *t-SNE* algorithm is a dimensionality reduction approach that is usually used to visualize better the underlying data. Since the cervigram images are in RGB format, we will flatten them first before *t-SNE* reduction. In Fig. 16, we report the data distribution using only two components. As indicated by the data scatter, separating the dataset cervigrams into two distinct clusters will very challenging.

Before starting the training process, the training subset must be augmented to ensure adequate training of our deep learning modules. The new dataset has the class distribution reported in Table 6. Note that the test subset is not augmented as it will lead to highly biased classification scores Goodfellow et al. (2016).

Finally, it should be noted that the augmented cervigram images are not pre-processed to mitigate the effects of specular reflection Saint-Pierre, Boisvert, Grimard, and Cheriet (2011).

Fig. 15. Samples from cancer-negative (left column) and cancer-positive class (right column).

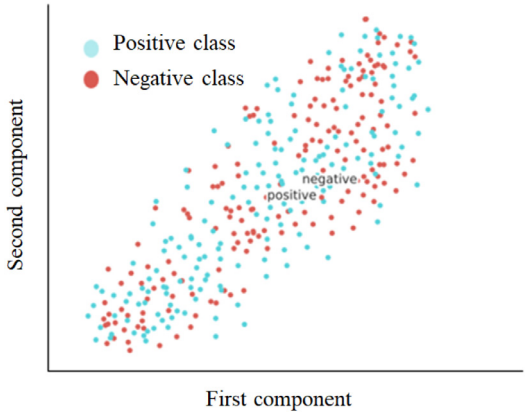

Fig. 16. t-SNE distribution for RGB representations of cervigram images.

4.3. Performance measures

4.3.1. RoI detection task

The efficiency of the cervix detection module is assessed using the IoU metric defined in Eq. (2). As mentioned earlier, the ground truth bounding boxes are manually labelled by a human expert Kaggle (2017b). We will report the average IoU scores for the N cervigram images in the validation subset as follows:

$$Avg\_IoU_p^t = \frac{1}{N}\sum_{i=1}^{N}\left(IoU_p^t\right)^{(i)} \tag{7}$$

where $\left(IoU_p^t\right)^{(i)}$ is the IoU score achieved at the detection of the cervix RoI in the ith cergvigram image.

4.3.2. Cervical cancer grade classification task

Binary classifiers are usually evaluated using various performance measures. However, most of these features are derived from four basic decision scores. These scores include:

1. True positive (TP): Number of positive samples correctly classified as positive.
2. True negative (TN): Number of negative samples correctly classified as negative.
3. False positive (FP): Number of negative samples incorrectly classified as positive.
4. False negative (FN): Number of positive samples incorrectly classified as negative.

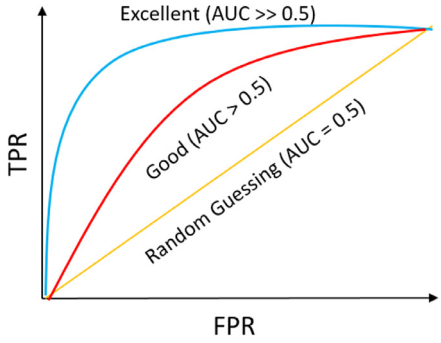

Fig. 17. Sample ROC curves of some binary classifiers.

Using the above, the true positive (TPR) and negative rates (TNR) are defined as follows:

$$TPR = \frac{TP}{TP+FN} \tag{8}$$

$$TNR = \frac{TN}{TN+FP} \tag{9}$$

Similarly, we can also define the false positive (FPR) and negative rates (FNR) as follows:

$$FPR = \frac{FP}{FP+TN} = 1 - TNR \tag{10}$$

$$FNR = \frac{FN}{FN+TP} = 1 - TPR \tag{11}$$

The TPR score is also known as the *sensitivity, recall* or *hit rate. Specificity, selectivity* and *true negative* are used to denote the TNR score in the literature. The precision or positive predictive value (PPV) of a classifier is given by:

$$Prec = PPV = \frac{TP}{TP+FP} \tag{12}$$

On the other hand, the accuracy is defined as:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

Since the classification scores, defined in Eqs. (8)–(13), depend on the decision threshold used, we will also report the performance results using the receiver operating characteristic curve (ROC). The ROC is evaluated using various decision thresholds. Finally, the ROC curve quality is usually quantified using the area-under-the-curve (AUC) score. The AUC score is a normalized score in the [0, 1] interval. Fig. 17 depicts the ROC curves produced by various binary classifiers.

4.4. Training and model validation

To ensure that the trained deep learning modules have been exposed to the training data, a 10-fold cross-validation strategy is adopted. The modules of the proposed deep learning pipeline are trained over 10 different epochs. At each epoch, 10% of the training data is used to assess the performance of the trained modules. As the training data has been balanced, stratified sampling is used to maintain data balance in each epoch. Then, performance scores averaged over the 10 epochs will be reported.

Table 7
System configuration of proposed cervix detection module.

Parameter	Description	Setting value
Batch Size	Number of images to train at a time	64
Subdivisions	The number of mini batches to load to memory	4
Momentum	The speed of moving the parameters in the optimization algorithm	0.9
Decay	The decay rate of the learning rate	0.0005
Learning rate	The distance of moving the parameters in the optimization algorithm	0.0001
Convolutional layers	The number of convolutional layers	22
Fully-connected layers	The number of dense layers	2
Labels	The number of classes	1

Table 8
Summary of detection accuracy and computational efficiency.

Method	IoU	Average time (*minutes*)
OBB Kim and Huang (2013)	0.736	20
ABB Song et al. (2015a)	0.699	4
IBB Malviya, Karri, Chatterjee, Manjunatha, and Ray (2012)	0.611	4
Proposed	**0.68**	**0.00367**

4.5. Results and discussion

The performance of the detection and classification modules is assessed and compared to existing state-of-the-art models. These existing models are designed using sophisticated hand-crafted features and machine learning models as described in Section 2. However, most of these techniques are data-driven and require considerable processing times.

4.5.1. Detection of cervix RoI

To speed up the training of the cervix detection module, a pre-trained model using the COCO dataset is trained Deng et al. (2009). Approximately 1200 images from the Intel&MobileODT annotated subset dataset are selected to train our proposed cervix RoI detection module. The remaining 300 annotated images are withheld for testing purposes. Table 7 gives a summary of our module configuration. To speed up training, images in each batch are further divided into small subsets as indicated by the *subdivisions* parameter. Also, the optimization algorithm is accelerated using the *momentum* value. As we are focusing on the cervix detection problem, only one object class is considered (**Class 0**). This assumption is also validated by the non-cancer nature of the cervigrams available in the Intel&MobileODT dataset Kaggle (2017a).

Table 8 summarizes the accuracy of the cervical RoI detection achieved by existing data-driven and proposed deep learning methods. In addition to their segmentation accuracy, the average computational performance is also reported. The OBB and ABB methods require the availability of the whole dataset for the prediction of the bounding box in each test image. Feature descriptors and extracted from 2000 and 939 cervigram images by the OBB and ABB methods, respectively. The average processing times are reported assuming cervigram datasets with the same numbers of images. Although the OBB method, proposed by Kim and Huang Kim and Huang (2013), achieves the best detection accuracy, it requires considerable processing power to extract the cervical RoI. On the other hand, our proposed RoI detection module, based on deep learning and automatically extracted features, attains similar detection accuracy in a real-time fashion. The superiority of the proposed detection method lies in its reliance on self-extracted features and the inception modules provided by the *Inception v4* and *YOLO* models. Moreover, the bounding box inference time is independent of the number of images in the dataset as the model has been trained to extract cervical RoI areas. In contrast, the data-driven OBB method generates sophisticated color and texture features for each image in the dataset and the test image Kim and Huang (2013). Approximately 2000 feature vectors are generated. Then, the test image feature descriptor is compared to those pertaining to all the dataset images. The top K matching images are retained for feature recalculation inside the predicted bounding boxes. Finally, the bounding box with the highest score is chosen as the bounding box of the test image. Similarly, the ABB method, attributed to Song et al. Song et al. (2015a), uses 939 labeled images to guide the detection of the bounding box in the test image Song et al. (2015a). First, PHOG descriptors are generated from the dataset and test images. Then, using a similarity measure, the top 20 matching images are selected. Finally, their bounding boxes are averaged to produce a bounding box for the test image.

It should be noted that the performance of the OBB and ABB methods highly depends on the dataset size. For instance, any reduction in the dataset size would cause a severe degradation in the performance of the OBB and ABB methods. This dependence on the dataset size makes these methods vulnerable to the overfitting problem Goodfellow et al. (2016). This problem can develop further as the datasets may contain blurry, deformed, rotated and transformed cervigram test images. In such cases, the detection process is likely to fail. On the other side, our detection method is robust to data variability as it employs a data augmentation strategy during the training process.

The IoU scores attained by our cervix detection model can be drastically improved by carrying out more training. The availability of larger annotated cervigram datasets will positively impact the overall performance of this model. Moreover, the confidence scores, provided by our detection model, represent a good measure for practitioners to assess its reliability. Fig. 18 reports the performance of the detection module and its confidence in properly detecting the cervical RoI areas. An interesting case is reported in the left side of the second row of Fig. 18 where our model has not only succeeded in detecting the cervix opening area but highlighted an area that looks like a second cevix opening. In this case, two bounding boxes with confidence scores of 70% and 50% are extracted by the model (see Eq. (1) for details). For automated detection, the bounding box with the highest score will be retained. In Fig. 19, more challenging cases are reported. In most of these cases, the cervical RoI is cluttered with blood or deformed tissues. However, the detection model can still extract the RoI areas accurately with confidence scores ranging from 90% to 70%.

4.5.2. Classification of cervical cancer grade

To highlight the efficiency of the automatic feature extraction and classification modules,**Model 1** and **Model 2** defined

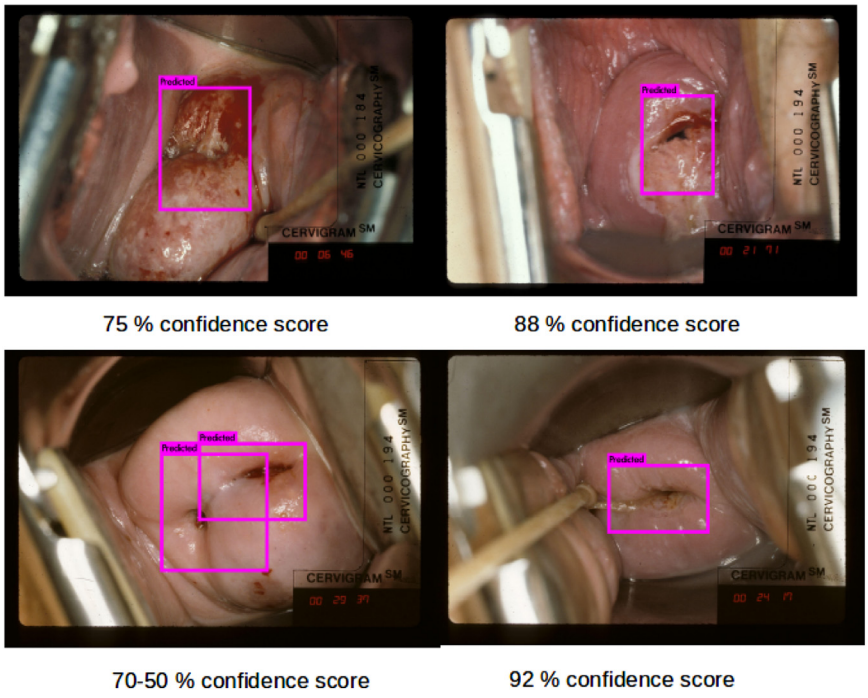

Fig. 18. Sample cervical RoIs with associated confidence scores.

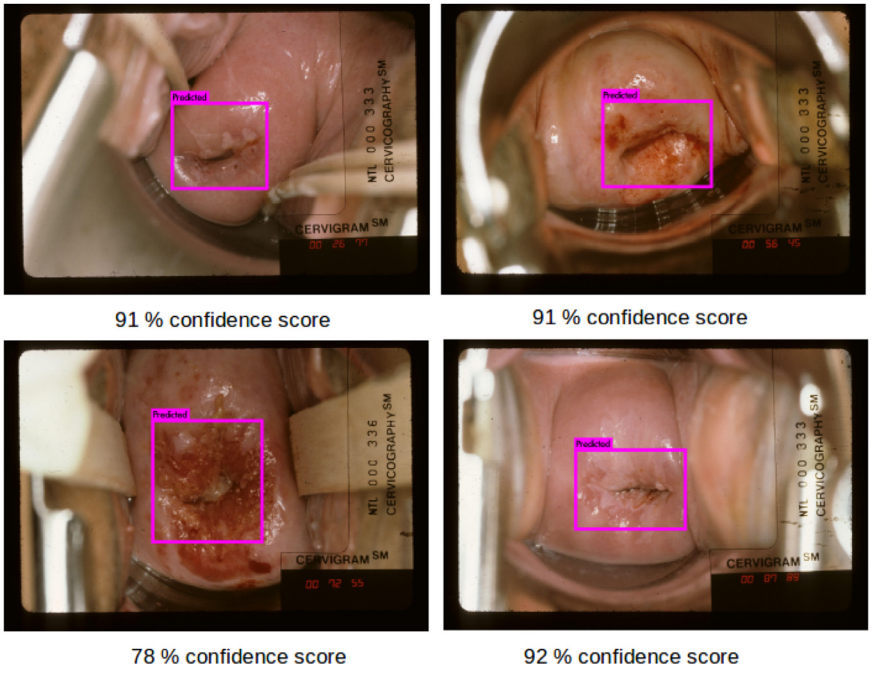

Fig. 19. Cluttered cervical RoI areas.

in Section 3.3, ANN-based classifiers using the hand-crafted features, described in Section 5, are also evaluated. As the extracted RoI areas can be varying in size, these image patches are rescaled to a standard size of 256×256 pixels prior to feature extraction and classification. Combined 2538-dimensional feature descriptors are extracted from each image in the *Guanacaste* project dataset. Then, two different ANN models are used to classify the resulting descriptors into cancer and non-cancer classes:

- **Model 3:** One hidden layer with 128 neurons followed by a ReLU activation layer. A 0.5 dropout layer is attached to the model to reduced the number of its parameters. The resulting output is processed through a sigmoid activation node.
- **Model 4:** Two hidden layers with 256 and 128 neurons, respectively. The remaining layers have the same structure as **Model 3** above.

Both classifiers are trained using an *adam* optimizer to minimize the cross-entropy cost function defined in Eq. (6).

Table 9
Classification performance scores using automatically-extracted (**Model 1** and **Model 2**) and hand-crafted features (**Model 3, Model 4**) and **Model 4**.

Model	Accuracy	Specificity	Sensitivity
Model 1	68.24 ± 9.74	77.43 ± 10.57	59.70 ± 12.08
Model 2	70.29 ± 8.57	68.33 ± 14.09	72.30 ± 13.85
Model 3	72.94 ± 6.94	76.80 ± 5.95	68.96 ± 9.96
Model 4	77.06 ± 7.06	77.97 ± 5.22	75.22 ± 11.28
Model 5	51.60 ± 4.63	51.69 ± 4.59	51.60 ± 4.63

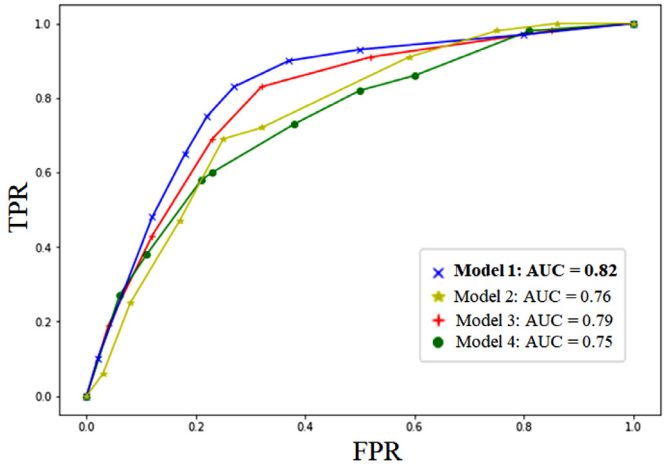

Fig. 20. ROC graph with AUC values for each model.

4.6. Training

Table 9 shows the performance results attained by the four models during the training phase. The averaged accuracy, specificity and sensitivity scores are calculated assuming a threshold level of **0.5**. Thanks to the automation of the feature extraction process, the proposed models, **Model 1** and **Model 2**), are able to classify a cervical image in 0.008 seconds. However, **Model 3** and **Model 4**) require 1.3 seconds to extract the compound feature descriptor from each cervigram. On the other hand **Model 5** takes around 4 seconds on each cervigram. The proposed models are **160** times faster that state-of-the-art existing models. **Model 1** achieves very high specificity scores which makes it very suitable for accurately diagnosing cancer-free cases. On the other hand, **Model 2** is recommended for cases with cancer suspicions as it attains very high sensitivity scores.

To reduce the effect of the detection threshold on the performance evaluation analysis, the ROC curves of the compared models in reported in Fig. 20. The proposed lightweight feature extraction and classifier, **Model 1**, attains the highest AUC score of **0.82** as it has been trained with a balanced dataset. The somehow deeper model, **Model 2**, achieved a lower AUC score as it was trained using the same dataset size while it should have been trained with more data as hinted by Fig. 6. Finally, the performance of the proposed models, **Model 1** and **Model 2**, can be further enhanced using training datasets with larger annotated cervigram images as these models are not data-driven. In fact, larger datasets will play favorably for our proposed models.

4.7. Failed cases

Before completing the performance analysis, we will shed the light on cervigram cases that caused the proposed detection and classification modules to fail.

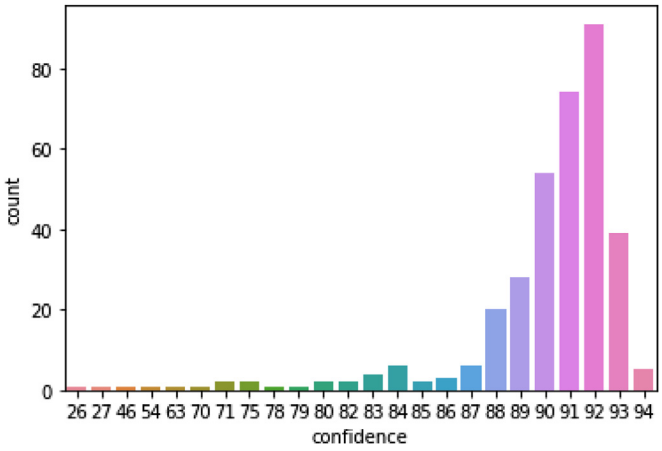

Fig. 21. Histogram of confidence scores over *Guanacaste* project dataset.

4.7.1. Cervix RoI detection

Fig. 21 provides a sketch on the distribution of the confidence scores attained by our detection model while detecting the cervical RoI area in all images in the *Guanacaste* project dataset. The historgram of the confidence scores is tailed towards the high range. More specifically, the reported scores are centered around an average score higher than 80%. In very few instances, the proposed detection model failed to detect the RoI area with reasonable confidence scores. Such low confidence scores are quite expected as *Guanacaste* project dataset contains cervigrams pertaining to the *CIN 4* cervical cancer grade. In such cases, the cervix region is occluded with blood stains or infected tissue as illustrated in In Fig. 22.

4.7.2. Classification

Figs. 23 and 24 provide samples of the resulting false positive and negatives, respectively. Two border cases of non-cancer, classified as cancer, are illustrated in Fig. 23. The non-cancer case, shown in the left side of Fig. 23, has the cervix region occluded in blood stains similar to cervical cancer cases. The missclassification can be easily rationalized. However, the right side of Fig. 23 represents a sample that is usually hard to diagnose as evidenced by the decision score of $p = .51$. Fig. 24 reports two border cases of cancer that are missclassified as non-cancer. The poor visual quality of the case, shown in the left side of Fig. 24, has caused the low decision score of $p = .13$. As the image quality degrades, the classifier is unable to learn any meaningful features pertaining to the cancer class. The right hand side of Fig. 24 shows a challenging case of cancer. This case looks very similar to the non-cancer one as the cervigram image does not provide enough clues about the presence of cancer. The classifier inability is characterized by the low decision score of $p = .34$.

Finally, classification failures call for a thorough investigation of the decision threshold, τ, that can be safely adopted to diagnose the presence of cervical cancer based on the analysis of the cervigram image. Using τ, the cancer classification decision is set as follows:

$$\text{decision} = \begin{cases} \text{Cancer}, & \text{if } p \geq \tau. \\ \text{Non-cancer}, & \text{otherwise}. \end{cases} \tag{14}$$

Setting an adequate value for τ is a challenging task since inadequate decision thresholds might lead to severe implications as in the case of biometric systems Telegrah (2009).

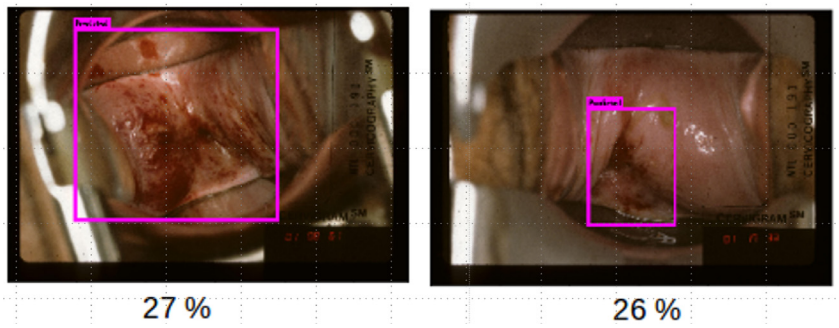

Fig. 22. Failed cases with confidence scores less than 30%.

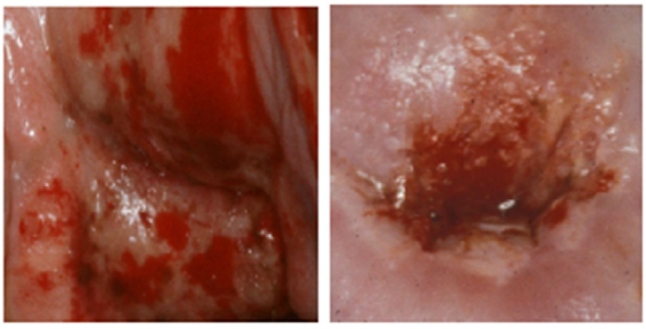

Fig. 23. False positive samples. $p = .65$ (left) and $p = .51$ (right).

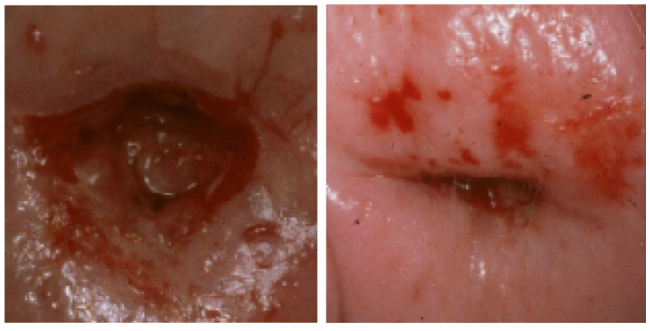

Fig. 24. False negative samples. $p = .13$ (left) and $p = .34$ (right).

5. Conclusions

This paper introduces a novel fully-automated deep learning pipeline for the detection of the cervix region and classification of cervical cancer. First, this pipeline consists of a detection module that detects the cervix region **1000** faster than state-of-the-art date-driven models and attains a detection accuracy of **0.68** in terms of intersection of union (IoU) measure. In addition, two lightweight versions of deep convolutional neural networks (CNNs) are proposed to classify cervical tumors. Self-extracted features, learned by the proposed CNN models, are used to classify cervical tumors. CNN-based classifiers outperform existing ones that are based on engineered features as they achieve AUC scores of **0.82** while classifying each image of the cervix region **20** times faster. Finally, the proposed deep learning pipeline is trained and evaluated using cervigram images pertaining to the *Guanacaste* project led by the American National Cancer Institute (NCI) and the Intel&MobileODT *Kaggle* challenge, respectively. The speed, accuracy and lightweight architecture of the proposed pipeline make it very suitable for deployment as a smart device application in less-developed countries. Such deployment will certainly reduce the fatalities due to cervical cancer. In future, we intend to enhance the perceptual quality of the cervigram images by reducing the effects of specular reflection and provide more accurate manual labeling of the cervical RoI.

6. Availability of data and materials

The cervical image datasets are obtained from:

1. NHI Guanacaste Project at: http://www.cse.lehigh.edu/idealab/cervitor.
2. Intel & MobileODT Cervical Cancer Screening at: https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data.

Software developed to support the conclusions of this article is available at:
https://github.com/lahouari2018/Cervical_Cancer_Classification_Using_Deep_Learning_Models.git.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Credit authorship contribution statement

Zaid Alyafeai: Software, Data curation, Visualization, Investigation, Validation. **Lahouari Ghouti:** Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing, Investigation, Validation, Supervision.

Acknowlgedgments

The authors would like to acknowledge the support provided by King Fahd University of Petroleum & Minerals (KFUPM).

Appendix A

Overview of hand-crafted Features

Given the importance of hand-crafted features and their use in the detection and classification modules of any cervical cancer diagnosis system, some of these features are discussed in this section. To boost the computation efficiency of the feature extraction process, RoI areas around the cervix are first extracted as indicated in Fig. 26-a. Then, the RoI region is down-sized to a typical size of 256×256 pixels Kim and Huang (2013). The down-sizing process is depicted in Fig. 26-b.

PLAB Descriptor

Kim and Huang indicated that good image descriptors can be extracted from the color and texture information of cervigrams Kim and Huang (2013). For instance, the PLAB feature descriptor is extracted from a pyramid of color histograms in the L*a*b* color

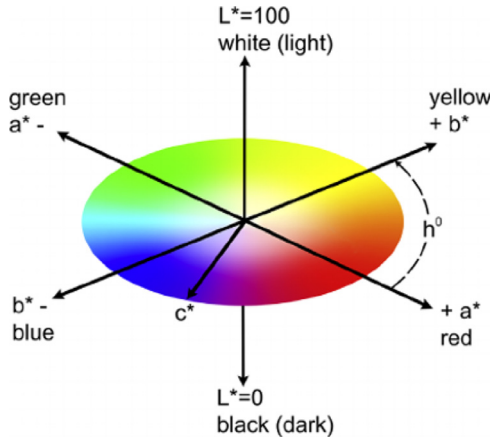

Fig. 25. L*a*b* color space.

space. Unlike in the standard red-green-blue (RGB) color space, a small color change in the L*a*b* space results in the same change in the visual appearance. The three-dimensional L*a*b* space represents the luminance (L* axis), the red-green dimension (a* axis) and the blue-yellow dimension (b* axis). The color-luminance separation is depicted in Fig. 25.

Since most of cervigram images are represented in the RGB color space, a color conversion is required before the feature extraction step. Fig. 26-c illustrates this process. Then, three *spatial* pyramids are extracted from the L*a*b* representation of the cervigram image. From each pyramid image, various blocks are considered as follows:

1. Level-1 pyramid: One block for the whole cervigram image.
2. Level-2 pyramid: The cervigram image is divided into four equal-size blocks.
3. Level-3 pyramid: Sixteen equal-size blocks are extracted from the cervigram image.

The resulting spatial pyramids are depicted in Fig. 26. To capture the image spatial layout and local color information, color histograms are estimated from all image regions in the 3 spatial pyramids. Then, each color histogram is quantized using 16 bins. Finally, the binned color histograms are concatenated to generate a PLAB descriptor with 1008 bins. The spatial pyramid at level 1 provides $3 \times 16 = 48$ bin values. Level 2 pyramid contributes with $3 \times 4 \times 16 = 192$ bin values. Finally, $3 \times 16 \times 16 = 768$ bins are obtained from the pyramid at level 3. Therefore, the PLAB descriptor is a 1008-dimensional vector. A typical PLAB descriptor vector is shown in Fig. 26.

PLBP Descriptor

The local binary pattern (LBP), developed by Ojala et al. Ojala, Pietikäinen, and Harwood (1996), was initially designed for grayscale images and then extended to color images and video sequences. Thanks to the LBP operator, a grayscale image are represented using an feature vector of integer numbers. This feature vector describes the small scale appearance of the image Ojala et al. (1996). Further processing of the feature vector is possible leading to more compact feature representations. Image textures are represented have two locally two complementary aspects including a pattern and its strength. Given a $S \times S$ image block, the LPB feature is generated for the block center pixel with respect to its $S^2 - 1$ neighboring pixels using:

$$LBP(x_c, y_c) = \sum_{p=0}^{S^2-1} \mathbb{1}(i_p - i_c \geq 0) \cdot 2^p \tag{15}$$

where i_c and i_p represent the intensities of the block center pixel and the $S^2 - 1$ neighboring pixels, respectively. The image location of the center pixel is given by the tuple (x_c, y_c). $\mathbb{1}(cond)$ defines an indicator function that takes a value of 1 when its argument, *cond* is true and 0 otherwise. Eq. (15) computes the value associated with the block center pixel as follows:

1. Each neighboring pixel is subtracted from the center pixel. Bit 1 is for a positive difference and 0 if the difference is negative.
2. Each resulting pixel is multiplied by 2^p where defines the bit neighboring pixel position rotationally and clockwise.
3. The sum of the $S^2 - 1$ scores above is used to represent the center pixel in the resulting LBP feature vector.

The generation process of the LBP feature vector is summarized in Fig. 27 for the center pixel of 3×3 image block.

The LBP descriptor can detect several texture primitives thanks to Eq. (15) as illustrated in Fig. 28. Spot and flat textures are shown in Fig. 28a-b. Fig. 28c-d illustrate textures with line ends and edges. Finally, a corner is represented by the LBP structure in Fig. 28-e.

It should be noted that the LBP descriptors extracted from 3×3, 5×5 and 9×9 describe radial image blocks with radius of 1, 2

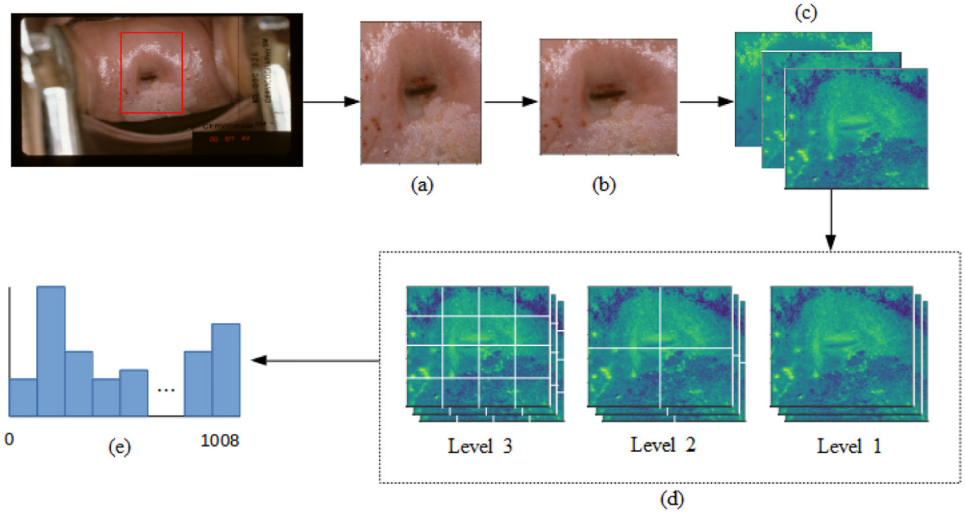

Fig. 26. Generation steps of PLAB descriptors.

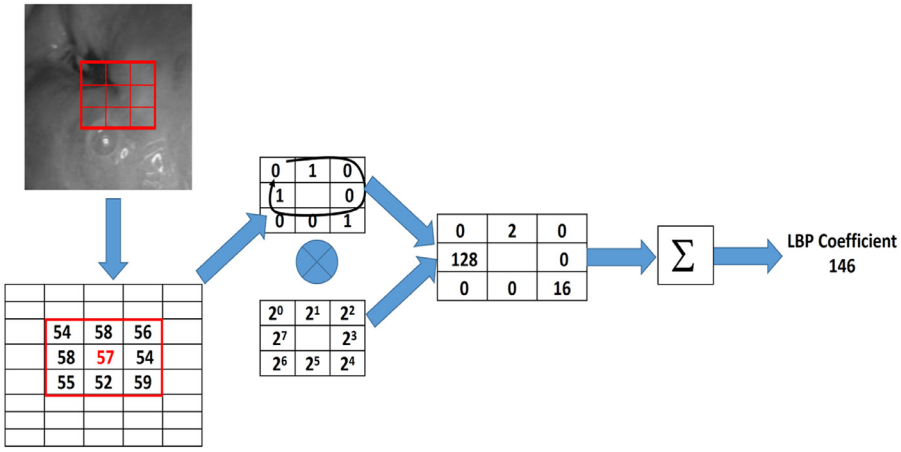

Fig. 27. Generation of LBP value for the center pixel of a 3×3 image block.

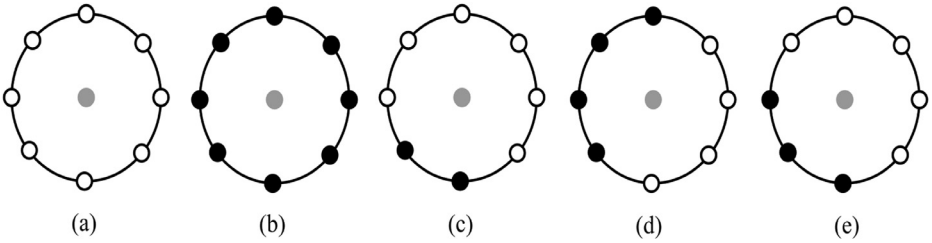

Fig. 28. Basic texture primitives detected by the LBP descriptor.

Fig. 29. Sample cervigram image (left). $LBP^{ri}_{8,1}$ descriptor (right).

Fig. 31. Sample cervigram image (left). Orientated gradient image (right).

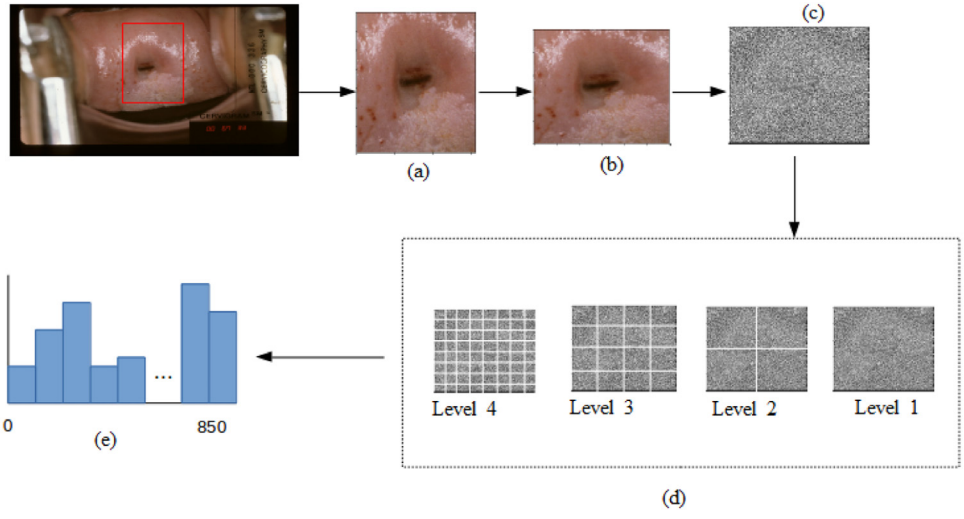

Fig. 30. Generation steps of PLBP descriptors.

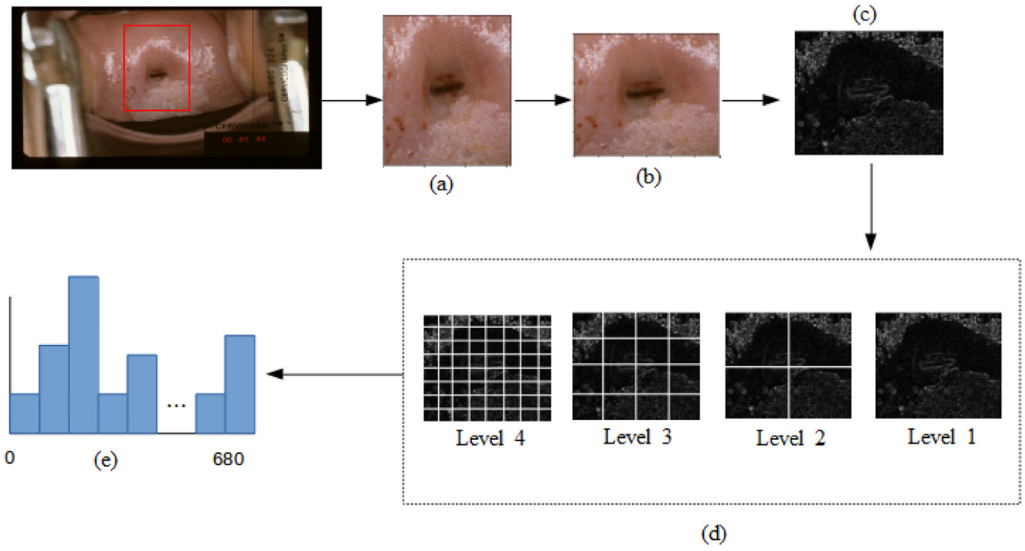

Fig. 32. Generation steps of PHOG descriptors.

and 3, respectively. However, the standard LBP is not invariant under image rotation Ojala et al. (1996). More specifically, rotating an image would cause the LBP descriptor to translate into a different location and rotate about its origin. To account for rotation invariance, the circular LBP descriptor is proposed. This descriptor, denoted by $LBP_{P,R}$, is generated using P equally-spaced pixels of radii R using:

$$LBP^{ri}_{P,R} = \min_i ROR(LBP_{P,R}, i), \ \ i = 0, \cdots, P-1 \tag{16}$$

where $ROR(bits, cir)$ defines the circular bitwise right rotation of the binary sequence $bits$ by $circ$ steps. A sample cervigram image with its corresponding $LBP^{ri}_{8,1}$ descriptor are depicted in Fig. 29 for non-overlapping image blocks with 8×8 size.

To generate the PLBP descriptor from cervigram images, Kim and Huang selected 8 neighboring pixels for each center pixel in the image block (i.e., $P = 8$ and $R = 1$) Kim and Huang (2013). Four pyramid levels are considered where the cervigram image is spatially split as follows:

1. Level-1 pyramid: One block for the whole cervigram image.
2. Level-2 pyramid: The cervigram image is divided into 4 equal-size blocks.
3. Level-3 pyramid: 16 equal-size blocks are extracted from the cervigram image.
4. Level-4 pyramid: The cervigram image is plit into 64 equal-size blocks.

Unlike the PLAB descriptor case, 10 bins are used to generate a final PLBP descriptor with 850 values. Fig. 30 describes the process used to generate the 850-dimensional PLBP descriptor from cervigram images.

PHOG Descriptor

Unlike the previous descriptors, PHOG is extracted from a gradient version of the cervigram image. Image gradients capture the edge information available in images. The unsigned orientation of these gradients (angles varied from 0^o to 180^o is depicted in Fig. 31.

On the other hand, the PHOG descriptor is extracted from 4 pyramid levels using the same block divisions (i.e., 1, 4, 16 and 64 blocks) as the PLAB descriptor. However, the HOG values are discretized using 8 bins only. Therefore, the resulting PHOG descriptors will consist of 680-dimensional vectors. The process for generating the 680-dimensional PHOG descriptor from cervigram images is illustrated in Fig. 32.

Composite Descriptor

To capture the color, texture and edge information available in the cervigram RoI, Kim and Huang concatenated the PLAB, PLBP and PHOG descriptors into a composite 2538-dimensional descriptor Kim and Huang (2013). Then, the cervical cancer classifier model is trained using these high-dimensional feature vectors.

References

Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16*(6), 641–647.

Adem, K., Kiliçarslan, S., & Cömert, O. (2019). Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification. *Expert Systems with Applications, 115*, 557–564.

AlMubarak, H. A., Stanley, J., Guo, P., Long, R., Antani, S., Thoma, G., … Stoecker, W. (2019). A hybrid deep learning and handcrafted feature approach for cervical cancer digital histology image classification. *International Journal of Healthcare Information Systems and Informatics (IJHISI), 14*(2), 66–87.

Almubarak, H. A., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., & Frazier, S. R. (2017). Convolutional neural network based localized classification of uterine cervical cancer digital histology images.. *Procedia computer science, 114*, 281–287.

Bamford, P., & Lovell, B. (1998). Unsupervised cell nucleus segmentation with active contours. *Signal processing, 71*(2), 203–213.

Bhargava, A., Gairola, P., Vyas, G., & Bhan, A. (2018). Computer aided diagnosis of cervical cancer using hog features and multi classifiers. In *Intelligent communication, control and devices* (pp. 1491–1502).

Chang, C.-W., Lin, M.-Y., Harn, H.-J., Harn, Y.-C., Chen, C.-H., Tsai, K.-H., & Hwang, C.-H. (2009). Automatic segmentation of abnormal cell nuclei from microscopic image analysis for cervical cancer screening. In *Nano/molecular medicine and engineering (nanomed), 2009 ieee international conference on* (pp. 77–80).

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). *Return of the devil in the details: delving deep into convolutional nets*. arXiv: 1405.3531.

Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J., & Chen, T.-J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics, 30*(1), 9–15.

Coleman, G. B., & Andrews, H. C. (1979). Image segmentation by clustering. *Proceedings of the IEEE, 67*(5), 773–785.

Cresce, R. P. D., & Lifshitz, M. S. (1991). Papnet cytological screening system. *Laboratory Medicine, 22*(4), 276–280.

Davies, E. R. (2012). *Computer and Machine Vision: Theory, Algorithms, Practicalities*. Academic Press.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on* (pp. 248–255).

Denny, L., Kuhn, L., Pollack, A., & Wright, T. C. (2002). Direct visual inspection for cervical cancer screening. *Cancer, 94*(6), 1699–1707.

Devi, M. A., Ravi, S., Vaishnavi, J., & Punitha, S. (2016). Classification of cervical cancer using artificial neural networks. *Procedia Computer Science, 89*, 465–472.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision, 59*(2), 167–181.

Fernandes, K., Cardoso, J. S., & Fernandes, J. (2017). Transfer learning with partial observability applied to cervical cancer screening. In *Iberian conference on pattern recognition and image analysis (ibpria 2017)* (pp. 243–250).

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning* (2nd). Springer, New York, USA..

Fukushima, K., & Miyake, S. (1982). Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern recognition, 15*(6), 455–469.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the ieee international conference on computer vision* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 580–587).

Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing* (4th). Pearson Education.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*.

Hartman, K. E., Hall, S. A., Nanda, K., Boggess, J. F., & Zoulnoun, D. (2002). Screening for cervical cancer: systematic evidence review. *US Department of Health and Human Services, Agency for Healthcare Research and Quality*.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Computer vision (iccv), 2017 ieee international conference on* (pp. 2980–2988).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Herrero, R., Schiffman, M. H., Bratti, C., Hildesheim, A., Balmaceda, I., Sherman, M. E., … Helgesen, K. (1997). Design and methods of a population-based natural history study of cervical neoplasia in a rural province of costa rica: the guanacaste project. *Revista Panamericana de Salud Pública, 1*, 362–375.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2017). Squeeze-and-excitation networks.

Hu, L., Bell, D., Antani, S., Xue, Z., Yu, K., Horning, M. P., … Befano, B., et al. (2019). An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI: Journal of the National Cancer Institute*.

Hubel, D. H., & Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *The Journal of physiology, 165*(3), 559–568.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456).

Jordan, J. (2009). *The cervix*. John Wiley & Sons.

Jordan, J., Singer, A., Jones, H., & Shafi, M. (2009). *The cervix*. John Wiley & Sons.

Kaggle (2017a). Intel&mobileodt cervical cancer screening dataset. Online; accessed 10-November-2018, https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data.

Kaggle (2017b). Manual annotation of intel&mobileodt cervical cancer screening dataset. Online; accessed 15-December-2018, https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/discussion/31565.

Kapur, J. N., Sahoo, P. K., & Wong, A. K. C. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing, 29*(3), 273–285.

Kim, E., & Huang, X. (2013). A data driven approach to cervigram image analysis and classification. In *Color medical image analysis* (pp. 1–13).

Koss, L. G. (1989). The papanicolaou test for cervical cancer detection: a triumph and a tragedy. *Jama, 261*(5), 737–743.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems.

Kudva, V., & Prasad, K. (2018). Pattern classification of images from acetic acid–based cervical cancer screening: a review. *Critical Reviews in Biomedical Engineering, 46*(2).

Kudva, V., Prasad, K., & Guruvare, S. (2018). Automation of detection of cervical cancer using convolutional neural networks. *Critical Reviewsâin Biomedical Engineering, 46*(2).

LaVigne, A. W., Triedman, S. A., Randall, T. C., Trimble, E. L., & Viswanathan, A. N. (2017). Cervical cancer in low and middle income countries: addressing barriers to radiotherapy delivery. *Gynecologic oncology reports, 22*, 16–20.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Li, F.-F., Karpathy, A., & Johnson, J. (2016). Cs 231n: Convolutional neural networks for visual recognition. Online; accessed 10-November-2018, http://cs231n.stanford.edu/.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research, 9*(Nov), 2579–2605.

Malviya, R., Karri, S. P. K., Chatterjee, J., Manjunatha, M., & Ray, A. K. (2012). Computer assisted cervical cytological nucleus localization. In *Ieee region 10 conference (tencon)* (pp. 1–5).

Mango, L. J. (1994). Computer-assisted cervical cancer screening using neural networks. *Cancer Letters, 77*(2–3), 155–162.

Mat-Isa, N. A., Mashor, M. Y., & Othman, N. H. (2005). Seeded region growing features extraction algorithm; its potential use in improving screening for cervical cancer. *International Journal of The Computer, the Internet and Management, 13*(1), 61–70.

Mayrand, M.-H., Duarte-Franco, E., Rodrigues, I., Walter, S. D., Hanley, J., Ferenczy, A., … Franco, E. L. (2007). Human papillomavirus dna versus papanicolaou screening tests for cervical cancer. *New England Journal of Medicine, 357*(16), 1579–1588.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition, 29*(1), 51–59.

Phoulady, H. A., Zhou, M., Goldgof, D. B., Hall, L. O., & Mouton, P. R. (2016). Automatic quantification and classification of cervical cancer via adaptive nucleus shape modeling. In *Ieee international conference on image processing (icip 2016)* (pp. 2658–2662).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Redmon, J., & Farhadi, A. (2017). Yolo 9000: Better, faster, stronger. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7263–7271).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Saint-Pierre, C.-A., Boisvert, J., Grimard, G., & Cheriet, F. (2011). Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications, 22*, 171–180.

Salembier, P., & Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing, 9*(4), 561–576.

van de Sande, K. E. A., Uijlings, J. R. R., Gevers, T., & Smeulders, A. W. M. (2011). Segmentation as selective search for object recognition. In *Ieee international conference on computer vision: 1* (pp. 1879–1886).

Sankaranarayanan, R., Basu, P., Wesley, R. S., Mahe, C., Keita, N., Mbalawa, C. C. G., … Nacoulma, M. (2004). Accuracy of visual screening for cervical neoplasia: Results from an IARC multicentre study in india and africa. *International Journal of Cancer, 110*(6), 907–913.

Schmidhuber, J. (2018). Who invented backpropagation?Online; accessed 10-November-2018, http://people.idsia.ch/~juergen/who-invented-backpropagation.html.

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv: 1409.1556.

Song, D., Kim, E., Huang, X., Patruno, J., Muñoz-Avila, H., Heflin, J., … Antani, S. (2015a). Multimodal entity coreference for cervical dysplasia diagnosis. *IEEE Transactions on Medical Imaging, 34*(1), 229–245.

Song, D., Kim, E., Huang, X., Patruno, J., Muñoz-Avila, H., Heflin, J., … Antani, S. K. (2015b). Multimodal entity coreference for cervical dysplasia diagnosis.. *IEEE Trans. Med. Imaging, 34*(1), 229–245.

Sornapudi, S., Stanley, R. J., Stoecker, W. V., Almubarak, H., Long, R., Antani, S., … Frazier, S. R. (2018). Deep learning nuclei detection in digitized histology images by superpixels. *Journal of pathology informatics, 9*.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning.. In *31st aaai conference on artificial intelligence (aaai-17): 4* (p. 12).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1–9).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2818–2826).

Szeliski, R. (2011). *Computer Vision: Algorithms and Applications*. Springer.

Tan, S. Y., & Tatsumura, Y. (2015). George papanicolaou (1883–1962): discoverer of the pap smear. *Singapore Medical Journal, 56*(10), 586–587.

Tang, J. (2010). A color image segmentation algorithm based on region growing. In *Proceedings of the 2nd international conference on computer engineering and technology: 6* (pp. 634–637).

Telegrah, T. (2009). Airport face scanners cannot tell the difference between Osama bin Laden and Winona Ryder. Online; accessed 10-November-2018, https://www.telegraph.co.uk/news/uknews/law-and-order/5110402/Airport-face-scanners-cannot-tell-the-difference-between-Osama-bin-Laden-and-Winona-Ryder.html.

Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2016). Global cancer incidence and mortality rates and trendsan update. *Cancer Epidemiology and Prevention Biomarkers, 25*(1), 16–27.

Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision, 104*(2), 154–171.

Wittet, S., Goltz, S., & Cody, A. (2015). *Progress in Cervical Cancer Prevention: The CCA Report Card 2015*. Cervical Cancer Action: A Global Coaliation to STOP Cervical Cancer.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Computer vision and pattern recognition (cvpr), 2017 ieee conference on* (pp. 5987–5995).

Xu, T., Zhang, H., Xin, C., Kim, E., Long, L. R., Xue, Z., … Huang, X. (2017a). Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern recognition, 63*, 468–475.

Xu, T., Zhang, H., Xin, C., Kim, E., Long, L. R., Xue, Z., … Huang, X. (2017b). Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognition, 63*, 468–475.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision., cham* (pp. 818–833).

Zhang, L., Lu, L., Nogues, I., Summers, R. M., Liu, S., & Yao, J. (2017). Deeppap: Deep convolutional networks for cervical cell classification. *IEEE journal of biomedical and health informatics, 21*(6), 1633–1643.