Automated Detection of Cervical Pre-Cancerous Lesions Using Regional-Based Convolutional Neural Network

Stephen Kiptoo[a]*, Lawrence Nderu[b], Leah Mutanu[c]

[a]*Student, Jomo Kenyatta University of Agriculture and Technology, P.O.Box 62,000-00200 Nairobi, Kenya*
[b]*Lecturer, Jomo Kenyatta University of Agriculture and Technology, P.O.Box 62,000-00200 Nairobi, Kenya*
[c]*Lecturer, United States International University USIU- Africa, P. O. Box 14634-00800 Nairobi, Kenya*
[a]*Email: Stephen.kiptoo@jkuat.ac.ke*
[b]*Email: lnderu@jkuat.ac.ke*
[c]*Email: lmutanu@usiu.ac.ke*

Abstract

The Cervical Colposcopy image is an image of woman's cervix taken with a digital colposcope after application of acetic acid. The captured cervical images must be understood for diagnosis, prognosis and treatment planning of the anomalies. This Cervix image understanding is generally performed by skilled medical professionals. However, the scarcity of human medical experts and the fatigue and rough estimate procedures involved with them limit the effectiveness of image understanding performed by skilled medical professionals. This paper, the model uses Regional Based Convolutional Neural Network (R-CNN) to effectively visualize of pre-cancerous lesions and to aid in diagnosis of the disease. The model was trained, on a dataset comprising of 10,383 cervical images samples. The datasets were derived from public dataset repositories. The training samples comprised of type class 1, 2 and 3 traits of cervical precancerous traits. The performance was evaluated using K-nearest - neighbor model over R-CNN. With an accuracy rate of 86%, this approach heralds a promising development in the detection of cervical precancerous lesions. This study findings established that the proposed model in provision of the better accuracy and misclassifications performance than various testing algorithms.

Keywords: Convolutional Neural Networks (CNN); CNN- Architecture; Cervical Colposcopy; Regional Based Convolutional Neural Network (R-CNN).

----

* Corresponding author.

1. Introduction

Cervical cancer is responsible for approximately 300,000 deaths worldwide, with 90% affecting women in low to middle income nations [1,2]. Health sectors, particularly within the management of chronic diseases, are facing great challenges with the rise of the aging population [3]. One of these challenges rises from several rural parts of the world, many women at high risk for cervical cancer are receiving treatment that will not work for them due to the position of their cervix. In Kenya for instance, cervical cancer has been established to be leading cause of death among and second most prevalent cancer affecting women, reporting about 5,250 new cervical cancer cases are diagnosed annually on women between 15 and 44 years [4]. The conventional cytological screening involves manual smearing and staining [5]. The complexity of the acquisition process for conventional cytology requires mobilizing expert teams to the field to evaluate [6] . Digital colposcopy is a very effective, non-invasive or pre-malignancy diagnostic tool used to review the pre-cancerous cases for cervical cancer [7] . In this procedure, the cervix is examined non-invasively using special methods designed by binocular stereomicroscope, where the abnormal cervical cell regions show aceto-white (AWE) lesion after application of 3-5% of acetic acid [8]. This is then biopsied using digital colposcopy guided for histopathological evaluation and confirmation [9]. Today, colposcopy today is considered reliable for detection and treatment of pre-cancerous lesions of the cervix, where women worldwide in low-resource settings are benefiting through programs where cancer is identified and treated or planned in a single visit [10]. Currently, a void in specialized image processing software which has the ability to process images acquired in colposcopy [11]. Several experts with different levels of expertise on Technology on various aspects through a competition organized by the analysis of vulnerable population, Intel and Mobile ODT on image quality assessment [12] and enhancement [13] of digital colposcopies', to the segmentation of cervical image anatomic regions of the [14] to the final diagnosis [15] and to automate the analysis of colposcopy images in order to support the medical decision process and to provide a data driven channel for communication of findings [16]. We are facing a possible turning point in the area, with the driving interest on the new advent of deep learning and neural networks [17]. Therefore, our work aimed at proposing a cervical pre-cancerous neural network architecture for detection of pre-cancerous lesions by making use of regional-based convolutional neural network.

2. Conceptual overview of convolutional neural network (convnet)

2.1   *Convolutional Layer*

This is a layer that has and accepts a volume of size parameters that consist of a set of learnable filters dimensions such as height, breath and number of channels that are essential in the convolution operation to derive the complex features from the edges of the input image [18]. These filters are small dimensional collections extending to a full depth of the input volume once matrix multiplication operations have been performed. This is done on a dot products amid logical regions of the inputs and the filters until it analyzes the whole width and replicates it till the pictorial is traversed [19]. This operation results in a 2-dimensional feature map that results to responses of that filter at every spatial dimension [20]. Instinctively, the network will learn filters that are activated when a type of visual feature such as an edge of some alignment is detected on the opening layer of a network. What results is an entire set of filters in each convolution layer with 12 filters, a 2-dimensional activation map [21]. These activation maps are stacked along the depth dimension resulting in an output volume

[22].

2.2 Rectified Linear Unit Layer(ReLU)

The activation function in neural network is responsible for transforming the summed weighted images input from the node into the activation of the node or output for that input [7]. Therefore, these ReLU activation functions are a piecewise linear function that will output the input directly if it is positive for positive type, otherwise, it will output zero for normal results. This layer performs pixel-wise operation and replaces all the negative values in the activation map by zero to get a rectified activation map [23]. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance faster and better the accuracy.

2.3   Fully Connected Layer

The cheapest way to possibly learn a non-linear function is through a fully connected layer. Outputs of the convolution layers are represented in space combinations of high-level features providing a fully expressive, low-dimensional and invariant attribute space [24]. All the neurons from the preceding layer be it pooling, convolution, or fully connected layers are taken and connected to each neuron within it. Each neuron receives weights that prioritize the most appropriate tag. Having transformed the input (any image) into an appropriate form of multi-level perceptron, the image is flattened into a column vector. The resulting flattened output is fed to a feed forward neural network as well as backpropagation which is applied to every repetition in the training [25]. The main goal of a fully connected layer is to accumulate the results of convolutions and the pooling processes in order to separate amid the dominating from certain low-level features and using them to categorize images into a basic classification.

2.4   Pooling Layer

The pooling layer uses height, width, depth, and stride that will control the amount of pixels and kernel size as it's parameters to reduce the spatial dimensions, but not depth, on a convolutional neural network model, to gain computation performance, less chance to overestimate or fit and to get some translation invariance [21]. The max-pooling operation, the layer operates on the input spatially resizing it with filter sizes of 2 by 2 dimensions which is the most common form of a pooling layer, two strides are applied across every depth along the width and the height ensuring that each max operation takes a maximum of four numbers all this is done while ensuring the depth pool layer dimensions remain unchanged [22]. Apart from max pooling operation, pooling units do perform other pooling events such as L2-norm pooling as well as average pooling. The creation of a brief version of features detected, is as a result of using pooling layers effectively generating scaled down or pooled feature maps [26]. To illustrate the usefulness of convnets in case of small alterations in the location of a feature in the input spotted by a convolutional layer, we resulted in a pooled feature map with the feature at a similar location [27].

2.5   Convolutional Neural Network Architecture

Convolutional neural networks (CNNs) recently have proved to be a remarkable success on neural natural language processing and computer vision, despite being biologically stimulated by the structure of mammals' visual cortex. [28] still followed up with same idea by adapting it to computer vision. CNNs typically comprise of different types of layers that can automatically acquire feature representation which is highly distinguishable without the need of hand-crafted features[29], suggested Alex-Net, a deep convnet architecture, comprising of seven hidden layers with millions of constraints that achieved a high-tech performance on the ImageNet dataset [30]. Conventional convolution operation requires a huge number of multiplications, which tends to increase the inference time and restricts the applicability of CNN to low memory and time constraint applications [31]. Many real-world applications, such as robots, healthcare, and mobile applications, perform the tasks that need to be conducted on computationally limited platforms in a timely manner. Hence different modifications in CNN are performed.

In addition, the image classification using convolutional neural network is a feed-forward process that following process [32]:

- First, feed the image dimensions (pixels) into the convolutional layer.
- Select the image features (parameters), apply strides and padding if it essential, and convolve the image using an activation map.
- Apply pooling to reduce the image dimensionality.
- Increase the convolutional layers until the model achieves the best accuracy desired.
- Then, it flattens the results and feeds into the last layer called a fully connected layer.
- Finally, produce the output using an activation function, which classifies the target image into the respective class.

The architecture outperformed other architectures outshining the second one which had an error rate of 26.2% while itself it had an error test of 15.3% [33]. Subsequently, the efforts saw the making of a very deep CNN architecture consisting of 16 layers which further significantly improved its accuracy. This pointed to a trend that showed that with every increasing depth of a CNN architecture there is a significant increase in its accuracy. This increase in depth resulted in a decrease in performance.
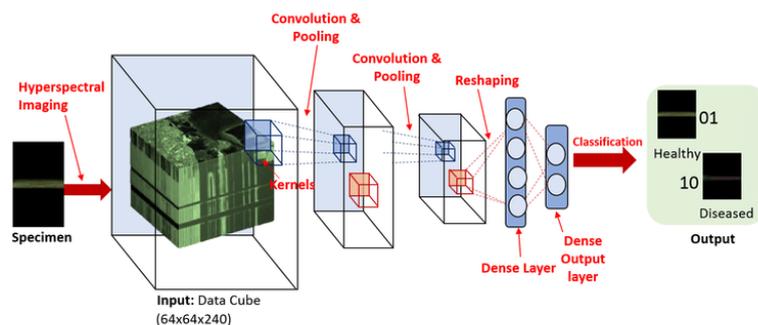

Figure 1: Architecture of Convolutional Neural network

Through research focusing on CNN reviews is still ongoing and has a room for significant improvement. Generally, observation occurrences between 2015 to 2020 shows a significant improvement in CNN performance. The capacity of a deep learning and CNN usually depends on its depth, and in a sense, an enriched variable set ranging from simple to more complex abstractions can aid in learning complex problems that arise.

However, the main setback faced by deep learning neural networks architectures is that of the diminishing the gradient. Previously, researchers attempted to eradicate this problem by connecting intermediate blocks of layers to auxiliary learners [34]. The emerging area of research was mainly the development of new connections to improve the convergence rate of deep CNN architectures. In this regard, different ideas such as information gating mechanism across multiple layers, skip connections, and cross-layer channel connectivity was introduced [35, 35]. To date, several improvements in CNN architectures have been reported. As this regard, the architectural advancement focus of research has been on designing of new layers that can scale up the network representation by exploiting feature-maps as well as processing input representation through adding more artificial inputs. Moreover, besides this, the focus is towards the design of the architectures type without affecting the performance to make CNN applicable at all device.

2.6   *Related work*

Author [36] developed a detection system that is able to differentiate between normal and cancerous cells by using colour intensity classification methods for automated extraction of multiple features of cytoplasm and nuclei from cervical cytology images are described (2011). The distinctive differences of colour intensity distributions between normal and cancerous cells have been successfully used to characterize cancerous cells. Author [37] discussed multiple feature extraction from Cervical cytology image by Gaussian mixture model. This method separates nucleus and cytoplasm from both single and multiple cellular Pap smear cervical cytology images using Gaussian mixture model (2014). Author [38] done a study towards developing an automatic image classification system by classifying Region of Interest (ROI). It is very difficult to identify abnormalities and using just the shape, size and gray-level information of a patient's cervix. Author [39] discussed detection of cervical cancer by using thresholding & watershed segmentation.An efficient and practical algorithm to detect the cervical cancerous area is presented in this paper. Magnetic resonance imaging scans of cervical cancer were acquired and its histogram image is segmented in addition with edge detection for extraction of tumorous area with exact edge and shape. The anatomic functional positioning as well as its effect on other cervix areas are also determined. MRI scans of cervical cancer patient are used and are stored in MATLAB in .jpg format. Images are then converted into a grey scale image based on intensity distribution of pixels in that image. Edge detection is applied on the image to find edges of tumour area. The hybrid segmentation is done on the sampled image i.e. combination of thresholding and watershed segmentation to extract the tumour infected region. The image is then processed using morphological image processing tools to detect the accurate dimensions of the tumour (2015).

3.   **Proposed methodology**

In this section an automated model was proposed using the reginal based convolution neural network architecture

capable of classifying the cervical precancerous lesions images.

3.1   Software

The neural network is designed in python 3.8.5 using TensorFlow and Keras deep learning libraries. Other python libraries like pandas, NumPy, seaborn, and matplotlib are also used in data preparation and visualization.

3.2   Data Collection

We obtained 10,383 de-identified images through available online dataset for training from Kaggle and UCI machine learning repository. And according to [40], these images were annotated as per the focus score output from a trained classifier and the annotations were quantized into types. Through this method, 10,383 cervical images were used as the training samples, 712 cervical images were used as test samples and 100 as validation samples.

3.3   Exploring the dataset

The design was done in the Jupyter notebook in anaconda environment. Initially, the dependent frameworks and libraries were installed into a local machine then imported into the Jupyter notebook editor. We used convolutional neural network (CNN) as well as transfer learning methods to evaluate pre-cancerous lesions, as per the anatomy of the cervix, and as by recommended screening guidelines [11]. Below are randomly selected images from Kaggle's cropped to include the ROI region, and subsequent steps of the process were performed within it, thus avoiding the confusing patterns and colors that occupy the rest of the image [41].

Figure 2: Few of Cervix Screened Images from UCI Repository

To address the confusion on assessment and enhancement of this medical cervical images' quality are often applicable to specific and require extensive domain knowledge.

3.4   Cervix Image Patch Extraction and Detection

A total of 10, 383 cervical precancerous lesions images were   represented in patches of 15×15 pixels with depth of 3 as shown in Fig. 2. Convolutional neural network trained a model to evaluate 9, 883 images which were used to train the CNN model, 712 for testing and 100 for validating as per the anatomy of the cervix; hence, the bounding boundaries information of the cervix region is crucial. Using the bounding box localization of the

cervix region, we can restrict the quality assessment to our area of interest (cervix) instead of including irrelevant regions [28].
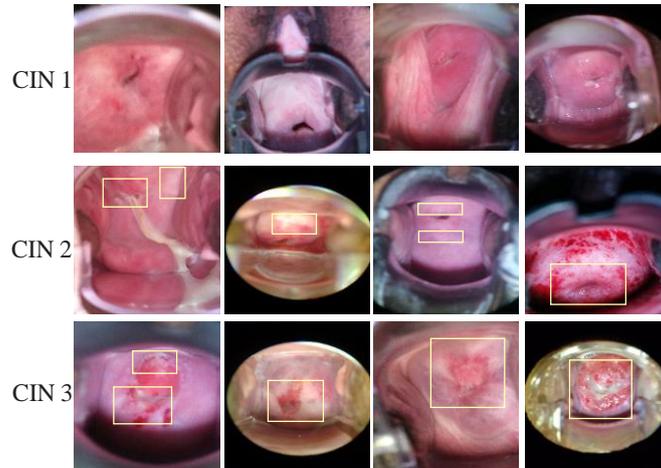

Figure 3: Examples of cervix detection images from pre-trained R-CNN

The goal for developing the training model using this proposed architecture was to achieve and obtain accuracy of the cervical precancerous lesions' coordinates for training purposes. The visual cortex of primates first receives input from the retinotopic area. Whereby, the lateral geniculate nucleus performs multi-scale high pass filtering and later contrast normalization [42]. We make the final decision by using the largest predicted score and the associated bounding annotation. This was manually verified, to define the automated detection of precancerous lesion to ensure accuracy of the trained and validated model in this research was achieved as illustrated in Fig. 3 on the separated findings.

3.5 *Technique Of Evaluating Neural Network Architectures*

Cross validation being a simple method running on multiple architectures while selecting the best performing architectures based on the authentication set was applied. In spite of this, only the best performing architectures is selected amidst architectures selected manually from a large number of choices. Exhaustive grid search has been considered the most effective strategy for hyper parameter optimization [12]. This is done by trying all the thinkable combinations manually for every stated hyper parameter. As in the works mentioned according to [43], the architecture is considerably shallow, with a couple of densely connected layers within three groups of convolutional pooling layers, and trained the architecture on a dataset with 485 images achieving 50% accuracy in recognizing three balanced classes [44]. Lately, random search has reported an improved result better than the grid search, by selecting hyper parameters randomly in a well-defined search space. Nevertheless, neither random nor grid search have been shown to use preceding valuations in identifying the succeeding set of hyper parameters for testing so as to better the anticipated architecture [45]. To calculate the accuracy of the model, the model must be compiled first. Compiling the model includes specifying the optimizer, loss function, and the metric evaluation method. Finally, the model uses accuracy to evaluate the model metrics. Below is the proposed model summary as illustrated in Fig. 4 this is the stage of checking the review of the model to confirm that everything is as expected

```
Model: "sequential_3"
_____
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_9 (Conv2D)            (None, 148, 148, 32)      896
_____
max_pooling2d_9 (MaxPooling2 (None, 74, 74, 32)        0
_____
conv2d_10 (Conv2D)           (None, 72, 72, 64)        18496
_____
max_pooling2d_10 (MaxPooling (None, 36, 36, 64)        0
_____
conv2d_11 (Conv2D)           (None, 34, 34, 128)       73856
_____
max_pooling2d_11 (MaxPooling (None, 17, 17, 128)       0
_____
conv2d_12 (Conv2D)           (None, 15, 15, 128)       147584
_____
max_pooling2d_12 (MaxPooling (None, 7, 7, 128)         0
_____
flatten_3 (Flatten)          (None, 6272)              0
_____
dense_5 (Dense)              (None, 512)               3211776
_____
dropout_1 (Dropout)          (None, 512)               0
_____
dense_6 (Dense)              (None, 1)                 513
=================================================================
Total params: 3,453,121
Trainable params: 3,453,121
Non-trainable params: 0
```

Figure 4: The algorithm of cervical pre-cancerous detection

Several technological researchers implement a hybrid of an already proposed architectures to improve deep CNN performance [44 ;46; 47; 48; 49; 50]. In 2018, they mainly on designing of generic blocks capable to be incorporated at any learning stage in CNN architecture, for improvement of the network representation [50]. In 2019, Khan and his colleagues introduced a new idea of channel boosting to up the performance of a CNN by learning distinct features as well as exploiting the already learned features through the concept of TL [51]. This has resulted in a probabilistic model built on preceding objective function evolutions Techniques that have implemented BO techniques include Sequential Model-based Algorithm Configuration (SMAC) which is grounded on random forests as well as Spearmint that uses a Gaussian process model [52]. Fig 1, describes the architecture of CNN, to achieve this evaluation accuracy.

4. Experiments and results

4.1 Model Configuration

We set up an algorithm using Keras environment, with other libraries to validate the accuracy of the performance. The entire dataset was split into a training set, Validation test and a testing set, in a ratio of 70:20:10. This was done automatically by initializing the weights and rescaling to 1. /255 units. A batch of 20 was used by a binary mode, with various optimizers such as Metrix to determine accuracies, cross entropy for loss checking and monitoring and RM-Prop at a Learning ration of 1e-5. We validated using 100 epochs using 50 validation steps.

4.2 Model Training

4.2.1 Experiment 1

The model was trained for a scaled down sized training set. This set up was also replicated for the testing samples. At each of these instances the number of validation samples were matched to the samples in the

training set. As observed from the results tabulated in Table I. In the first two instances the models were trained for 20 epochs.

Table 1: Model accuracy and loss outcome for the training samples outcome index I

Epochs	Loss	Accuracy	Val loss	Val Acc
20	0.4026	**0.2625**	0.4343	0.2550
20	0.3609	**0.3075**	0.4169	0.2640
20	0.3203	**0.3216**	0.3667	0.3170
20	0.2799	**0.3515**	0.4179	0.2530
20	0.2518	**0.3655**	0.3572	0.3120
20	0.2242	**0.3885**	0.3621	0.3120
20	0.1921	**0.4095**	0.3687	0.3180
20	0.1688	**0.4320**	0.3479	0.3370
20	0.1365	**0.4615**	0.3517	0.3260
20	0.1254	**0.4625**	0.3560	0.3120

Training of the model beyond 20 epochs, saw it suffer the intense overfitting as illustrated in Figure 5. which in turn reversed the gains that had been learned by the model. As such to prevent the model from negative gain the maximum number of epochs was capped to 20.
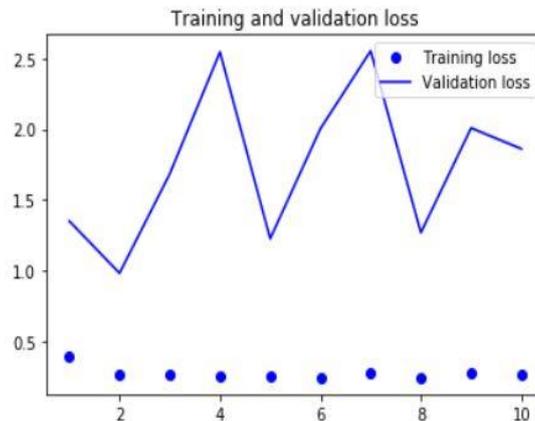

Figure 5: Model training and validation loss for training sample outcomes index II

4.2.2   Experiment 2

Further, training of such a model would result in the model gaining further characteristics which would negatively impact the model resulting in overfitting. This is captured in Figure 5. For each of the remaining instances the models were set for training over 10 epochs as illustrated in Table 2.

Table 2: Model accuracy and loss outcome for the training samples outcome index II

Epochs	Loss	Accuracy	Val loss	Val Acc
10	0.4003	**0.9190**	1.3499	0.5000
10	0.2763	**0.9521**	0.9821	0.5000
10	0.2652	**0.9520**	0.6832	0.5000
10	0.2551	**0.9520**	2.5443	0.5000
10	0.2528	**0.9535**	1.2257	0.5000
10	0.2425	**0.9566**	2.0025	0.5000
10	0.2734	**0.9515**	2.5517	0.5000
10	0.2469	**0.9520**	1.2679	0.5000
10	0.2736	**0.9480**	2.0071	0.5000
10	0.2720	**0.9500**	1.8601	0.5000
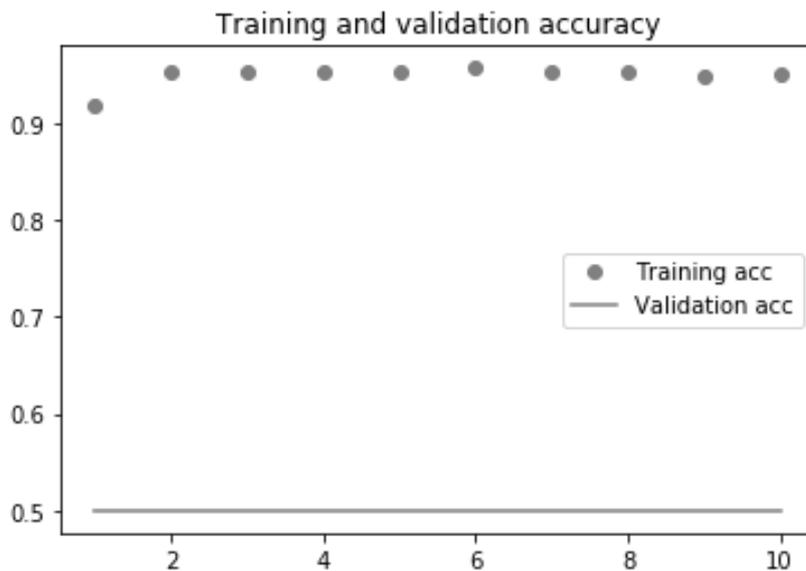
Figure 6: Model training and validation accuracies for training sample outcomes index V

4.2.3   Experiment 3

By turning various parameter, sampling and epochs the model accuracies trained for 20 epochs appear to have the lowest result between 30%-40%. The increase observed could be attributed to the increase in training as well as testing samples. This low result observed is when a comparison is made to the models trained for 10 epochs which post an accuracy result between 48%-86%. Gradually increasing the training samples together with the testing samples results to a steady increase in terms of the accuracy.

Table 3: Training Samples Outcomes

Index	Epochs	Training Samples	Testing Samples	Accuracy
I	20	500	75	0.3365
II	20	1000	200	0.3627
III	10	6000	200	0.4814
IV	10	2000	1000	0.7366
V	10	10383	712	0.8578

Initially, the model trained for 500 training samples post a 33% accuracy compared to the model trained for 10,383 samples which posted an 86% accuracy. Training accuracies for models for 20 epochs appear to have a steady increase in the accuracy and validation accuracies. This is partly because of the low number of training samples as well as testing samples. The result point to a low outcome ranging between 30% and 40%. On increasing the training samples to 6000 figures with about 200 in training samples the accuracy in predictions of the trained model increases slightly to 48%. Observations from the training and validation accuracy diagram in Table 2, there appears to be a steady increases terms of training accuracies then a decline in validation accuracies. This appears to suggest an increase in training samples with merely enough testing samples will affect the accuracy of the trained model. On increasing the samples to 2,000 and having the test samples at 1,000, the prediction accuracy of the trained model rose to 73.66%. Further increase of training samples to 10,383 raises up the performance of the model to 85.78%. as illustrated in the Table 3. The accuracies from a model utilizing 500 training samples and 75 testing samples were used in table 1. As would be expected for every model, on further training a model is expected to gain as successive and higher order convolution features are obtained. There is a gradual but systematic increase in terms of the accuracy outcomes as observed. As such the model gains marginally but does not get enough samples that would enable it learn the most significant features of would be a cervical pre-cancerous case. Table 3. documents the accuracies obtained from a model case with 10,383 training samples and 712 test samples. The model accuracies point to a significant increase in terms of training accuracies compared to the previous case. As observed from the documented accuracies there is an increase followed by a decline then an increase.

4.2.4   *Experiment 4*

Table 5: Precision, Recall and F1 Score of KNN Classifier

	precision	recall	f1-score	support
1	0.18	0.34	0.24	179
2	0.46	0.63	0.53	869
3	0.51	0.26	0.35	985
accuracy			0.43	2033
macro avg	0.38	0.41	0.37	2033
weighted avg	0.46	0.43	0.42	2033

Confusion matrix was introduced to further describe the performance of images classification of our model on a set of test data for which the true values are known. We experiment using precision, recall, accuracy, F1 score and ROC. Below was the resulted based on the three classes of the images we used.

These comparison outcome shows that the unreliability of KNNs compared to CNNs. This was based on the sample sizes used to train the algorithm for each class that determines the output in terms of accuracy. The test was based on the three types 1,2,3 that was 2000 testing samples. The accuracy shown here was 43% which demonstrates that the unreliability of KNNs compared to CNNs. Therefore, the accuracy was based on the individual image types and the availability of sufficient training samples. The confusion matrix was generated after using the following training image samples which comprised of 714 for type 1, 3477 for type 2, 1941 for type 3. The testing image samples were comprising of 679 for type 1, 869 for type 2, 485 for type 3. This gives a total of 6132 training samples and 2033 (60,549 and 261) testing samples as shown in Figure 7.

Confusion matrix without normalization
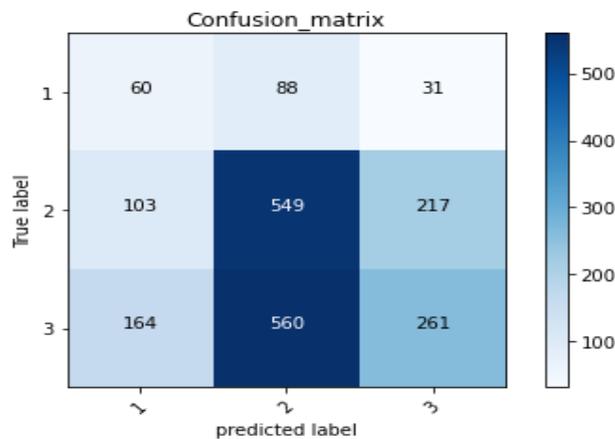
[60 88 31]

[103 549 217]

[164 560 261]

Figure 7: Values outcome for Confusion Matrix with normalization for test samples.

Normalized confusion matrix

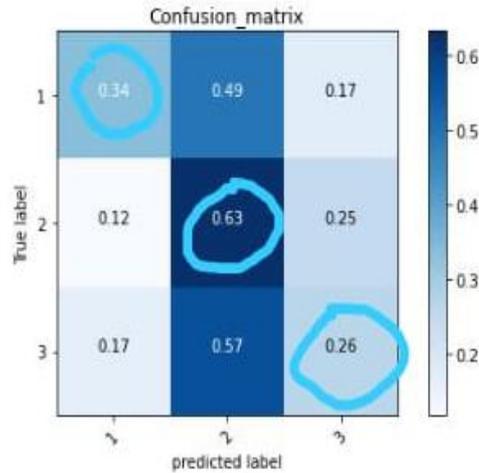[0.34 0.49 0.17]

[0.12 0.63 0.24]

[0.17 0.57 0.26]

Figure 8: Normalized confusion matrix

The highlighted values in Figure 8 shows that they are correctly classified in percentages. Depending on the sample's sizes used to train the algorithm for each class type that determines the output accuracies. The Comparison is also made on the AUC and ROC of a model to decide which model to implement on the given dataset for correct prediction. Therefore, 714 samples yield 0.34%, 3477 samples yield 0.63%, and 1941 samples yields 0.26% accuracies. This also shows a high value predict for the second label, which confirms that in type class 2 we had 3477 testing samples, for such since we did not have the successive and high-level convolutions matrixes, we had 0.49(49%) and 0.57(57%) misclassification rates for the samples type class 1 and type class 3 as highlighted in Figure 9. The results prove that CNN is has the best prediction of accuracy than KNN, hence suggest to adopt CNN in image classification and detection of precancerous lesions.
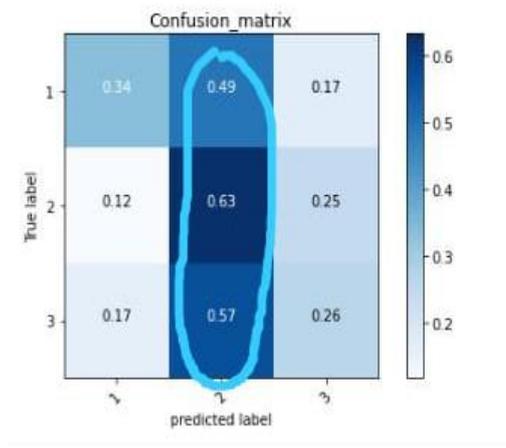

Figure 9: Misclassification Output using KNN

4.3 Strength and Limitation

This is the research reported better performance to the best of our knowledge. However, our study was not without limitations. There was the likelihood of overfitting the model during analysis due to a small number of

testing samples. However, this was minimized by restricting the selection of variables to only those identified as relevant based on experience and previous anatomy literature.

4.4   Discussions

In this study, deep learning, the Keras neural network and TensorFlow libraries were found to be the best program to be used for automated detection of precancerous lesions. The present study investigated whether neural networks could be applied to the detect any lesions from the cervical images using colposcopy. Using a deep learning approach called Regional Based CNN with extensive image augmentation based on a pre-trained model; we trained and validated an image analyzer that performs "automated visual evaluation" of the cervix. From the analysis, we noted that training model requires sufficient samples size. Therefore, several types of cervical images are increasingly available through Internet via public repositories, and inexpensive high-end smartphones or digital colposcopies are readily available for the general researchers, facilitating the uploading and sharing of this information, like as pictures for data and image processing. Currently, machine learning and statistical analysis can be performed using high-performance personal computers which are affordable for individuals, where there is limited volume of information. In addition, deep learning and neural networks technologies are becoming more accessible for research individuals and interested corporations. For instances, the Google software library for machine learning, TensorFlow, which was released under an open-source license in 2015 [53]. Based on these emerging trends, the present study aimed to apply deep learning neural networks to gynecological clinical practice. Unlike other several research studies [18,54] performed based on traditional machine learning approaches like the SVM, Reference [15] showed approximate 50% validation classification accuracy. However, recently most researches have been focused on improving the recognition accuracy of models. Very little studies have been demonstrated on recurrent architectures using convolutional neural networks. This strategy is an exceptional for context modeling from inputs, while this result in itself is not satisfactory, proposing that deep learning techniques has the potential to detect any lesion from cervix images from colposcopy. Therefore, this research relied on deep learning approaches to leverage on the systematic and successive higher-level feature extraction. The training was performed on numerous sets of training cervix images samples while documenting the outcomes at each time. However, at each of the instances the following sets of parameters was kept constant; the training batch was set to 20 with the model being trained for 10 epochs with 100 steps being set for each epoch. The architecture used consists of several blocks of convolutional layers followed by a max pooling layer. In the second to last layer of the structure global max-pooling is used, which is followed by a soft-max layer at the end [55]. This architecture demonstrated the state-of the-art accuracy for object classification at the time [56, 57]. This architecture uses a combination of two popular techniques, CNN and LSTM. The extraction is performed to identify how features vary with respect to time. This proposed model shows better performance for visual images analysis. In stark, contrast to these research findings, other previous studies by [58] shows, the training set accuracy using deep CNN was 80.1% and according to [18] used a randomly selected 345 images, and reported an accuracy of 83%, by resizing the images to 120*120pixels and passing it to LeNet CNN. Our proposed model using R-CNN recorded 85.75% which is slightly higher on training samples of 10,383 improved up the performance of the model. Despite the high accuracy rating, Convolutional Neural Networks shows to have a better performance than K-nearest-neighbor model. The detection of cervical precancerous traits is not always a straight forward feat. Low quality, low resolution

images, presence of foreign such as IUD or hair makes the detection algorithm susceptible to erroneous classifications. Perhaps to ensure more accuracy rating careful selection and prepping of the images should be undertaken.

4.5   Recommendations and Future Work

This paper proposes an automated detection of cervical pre-cancerous lesion using convolutional neural networks. Unlike previous methods which were based on visual impression which the health provider can miss the diagnosis during visualization with necked eyes and also cytology which is read microscopically for features extractions and also is very expensive in developing counties.   The Proposed model yields the higher performance, as compared to previous state of the art approaches in terms of accuracy, precision, recall, F1 score and more unites. This anticipates the segmentation free, high accuracy detection of the lesion, and classification of this model is a promising approach during screening using visual inspection with acetic acid. In the future, the impact of the model on field of viewing cervical images with overlapping cells is to be analyzed to avoid the misclassifications of images. Furthermore, Deep learning-based cervical images segregation still need more exploration for high accuracy. In addition, the number of samples and epochs tested outlines that there is need to establish a parameter configurations guideline which should be followed while configuring the parameters used by R-CNN. Some of our recommendations include:

- Number of training, testing and validating size: Samples sizes used to train the algorithms for each class determines the accuracies, therefore accuracies depend on individual classes and then the availabilities of sufficient training samples.
- Misclassification: To ensure the model captures all features which can misclassify the cervical images. We did not have successive and high-level convolutional matrixes as in CNN which lead to report of 0.49(49%) and 0.57(57%) misclassification rates for samples.
- Number of epochs: This parameter relates to the intervals during training. This tends to ensure the model get all feature for a model training. We recommend more epochs to improve the accuracies of the model.
- Optimization size: These supports the adaptive learning rate methods; these values depend on application and dataset; We recommend good values that can be for default are 50 and a millionth.
- Other Traits: These are other features which can lead to misclassifications of the cervical images. We do recommend considerer the presence of foreign features during classifications.

The deep learning algorithm could accurately detect any abnormal conditions that would require immediate treatment as per the cancer guidelines and which would not. This research, used and trained model using different cervical images from various regions, and suggested that method of detection may be useful in areas with limited resources, due to a shortage of healthcare workers in these areas, the AI method may make up for this fact and allow for accurate and quick diagnosis. For instance, a mobile phone camera can be used to diagnose for precancerous cervical cancer during a single visit. Furthermore, directions for this research team include training the algorithm further on images of cervical precancer and normal cervical tissue samples from all over the world using different types of cameras and imaging tools. There are small variations in the

appearance of the cervix due to presence of foreign traits. Then, they would allow this to be an open access resource with the ultimate goal of really reducing cervical cancer globally.

5. Conclusion

Cervical cancers are screened manually by using the visual inspection with acetic acid to detect any lesion that can lead to cancer. With significant strides having been made in diagnosis and detection of these ailments, the same has been replicated in the detection of the precancerous lesions. In this paper, the model proposed R-CNN model demonstrates its ability to detect the presence of cervical pre-cancerous traits and could highly aid in diagnosis of imminent progression of cervical precancerous traits. To ensure higher classification accuracies the proposed R-CNN model made use of significantly sufficient training as well as testing samples. In order to make this efficient and effective, various computerized methods are discussed to enhance better results.

Acknowledgments

References

[1]. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[2]. F. Islami, L. A. Torre, J. M. Drope, E. M. Ward, and A. Jemal, "Global cancer in women: Cancer control priorities," Cancer Epidemiol. Biomarkers Prev., vol. 26, no. 4, pp. 458–470, 2017, doi: 10.1158/1055-9965.EPI-16-0871.

[3]. World Health Organization, GUIDE TO CANCER - Guide to cancer early diagnosis. 2017.

[4]. "Kenya Cancer Statistics & National Strategies," Kenyan Network of Cancer Organizations, Feb. 18, 2013. https://kenyacancernetwork.wordpress.com/kenya-cancer-facts/ (accessed Aug. 13, 2020).

[5]. "(1) (PDF) Screening for Cervical Cancer Using Automated Analysis of PAP-Smears," ResearchGate. https://www.researchgate.net/publication/261923436_Screening_for_Cervical_Cancer_Using_Automated_Analysis_of_PAP-Smears (accessed Aug. 14, 2020).

[6]. "Wei et al. - 2017 - Cervical cancer histology image identification met.pdf." .

[7]. X. Q. Zhang and S. G. Zhao, "Cervical image classification based on image segmentation preprocessing and a CapsNet network model," Int. J. Imaging Syst. Technol., vol. 29, no. 1, pp. 19–28, 2019, doi: 10.1002/ima.22291.

[8]. "National-Cancer-Screening-Guidelines-2018.pdf." Accessed: Aug. 13, 2020. [Online]. Available: https://www.health.go.ke/wp-content/uploads/2019/02/National-Cancer-Screening-Guidelines-2018.pdf.

[9].   L. Wei, Q. Gan, and T. Ji, "Cervical cancer histology image identification method based on texture and lesion area features," Comput. Assist. Surg., vol. 22, no. sup1, pp. 186–199, Oct. 2017, doi: 10.1080/24699322.2017.1389397.

[10].   "Intel & MobileODT Cervical Cancer Screening." https://kaggle.com/c/intel-mobileodt-cervical-cancer-screening (accessed Aug. 14, 2020).

[11].   World Health Organization, Ed., WHO guidelines for screening and treatment of precancerous lesions for cervical cancer prevention. Geneva: World Health Organization, 2013.

[12].   K. Fernandes, D. Chicco, J. S. Cardoso, and J. Fernandes, "Supervised deep learning embeddings for the prediction of cervical cancer diagnosis," PeerJ Comput. Sci., vol. 4, p. e154, May 2018, doi: 10.7717/peerj-cs.154.

[13].   "Maini and Aggarwal - 2010 - A Comprehensive Review of Image Enhancement Techni.pdf." .

[14].   "(1) (PDF) A Novel Analysis of Clinical Data and Image Processing Algorithms in Detection of Cervical Cancer," ResearchGate. https://www.researchgate.net/publication/277667329_A_Novel_Analysis_of_Clinical_Data_and_Image_Processing_Algorithms_in_Detection_of_Cervical_Cancer (accessed Aug. 14, 2020).

[15].   M. Sato et al., "Application of deep learning to the classification of images from colposcopy," Oncol. Lett., Jan. 2018, doi: 10.3892/ol.2018.7762.

[16].   N. Muinga et al., "Digital health Systems in Kenyan Public Hospitals: a mixed-methods survey," BMC Med. Inform. Decis. Mak., vol. 20, no. 1, p. 2, Dec. 2020, doi: 10.1186/s12911-019-1005-7.

[17].   M. Nielsen, "Neural networks and deep learning." 2019.

[18].   A. Mittal and M. Juneja, "Cervix Cancer Classification using Colposcopy Images by Deep Learning Method," Int. J. Eng. Technol. Sci. Res., vol. 5, no. 3, pp. 426–432, 2018.

[19].   C. Data Science, "An Intuitive Explanation of Convolutional Neural Networks – the data science blog_." 2017.

[20].   R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Imaging, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[21].   H. A. Almubarak et al., "Convolutional Neural Network Based Localized Classification of Uterine Cervical Cancer Digital Histology Images.," Procedia Comput. Sci., vol. 114, pp. 281–287, 2017, doi: 10.1016/j.procs.2017.09.044.

[22].   Guillaume Berger, "CS231n Convolutional Neural Networks for Visual Recognition." 2016.

[23].   J. Brownlee, "A Gentle Introduction to the Rectified Linear Unit (ReLU)," Machine Learning Mastery, Jan. 08, 2019. https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ (accessed Aug. 14, 2020).

[24].   "CS231n Convolutional Neural Networks for Visual Recognition." https://cs231n.github.io/convolutional-networks/ (accessed Aug. 14, 2020).

[25].   "Convolutional Neural Network. In this article, we will see what are… | by Arunava | Towards Data Science." https://towardsdatascience.com/convolutional-neural-network-17fb77e76c05 (accessed Aug. 14, 2020).

[26].   J. Brownlee, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks," Machine

Learning Mastery, Apr. 21, 2019. https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/ (accessed Aug. 14, 2020).

[27]. "Litjens et al. - 2017 - A Survey on Deep Learning in Medical Image Analysi.pdf." .

[28]. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," Adv. Neural Inf. Process. Syst., vol. 2, no. January, pp. 1799–1807, 2014.

[29]. M. R. Minar and J. Naher, "Recent Advances in Deep Learning: An Overview," vol. 2006, pp. 1–31, 2018, doi: 10.13140/RG.2.2.24831.10403.

[30]. M. Wu, C. Yan, H. Liu, Q. Liu, and Y. Yin, "Automatic classification of cervical cancer from cytological images by using convolutional neural network," Biosci. Rep., vol. 38, no. 6, pp. 1–9, 2018, doi: 10.1042/BSR20181769.

[31]. "Detecting Driver Drowsiness in Real Time Through Deep Learning Based Object Detection," springerprofessional.de.   https://www.springerprofessional.de/en/detecting-driver-drowsiness-in-real-time-through-deep-learning-b/16772484 (accessed Aug. 14, 2020).

[32]. Prabhu, "Understanding of Convolutional Neural Network (CNN) — Deep Learning," Medium, Nov. 21, 2019. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148 (accessed Aug. 27, 2020).

[33]. G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[34]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," ArXiv151200567 Cs, Dec. 2015, Accessed: Aug. 15, 2020. [Online]. Available: http://arxiv.org/abs/1512.00567.

[35]. R. Barth, J. Hemming, and E. J. Van Henten, "Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation," Comput. Electron. Agric., vol. 173, p. 105378, Jun. 2020, doi: 10.1016/j.compag.2020.105378.

[36]. E. Supriyanto, N. A. M. Pista, L. H. Ismail, B. Rosidi, and T. L. Mengko, "Automatic Detection System of Cervical Cancer Cells Using Color Intensity Classification," p. 5.

[37]. G. K. Lakshmi and K. Krishnaveni, "Multiple Feature Extraction from Cervical Cytology Images by Gaussian Mixture Model," in 2014 World Congress on Computing and Communication Technologies, Feb. 2014, pp. 309–311, doi: 10.1109/WCCCT.2014.89.

[38]. S. Roy, R. P. Chauhan, and G. K. Verma, "Cervical cancer detection from MR images based on multiresolution wavelet analysis," in 2016 11th International Conference on Industrial and Information Systems (ICIIS), Dec. 2016, pp. 788–793, doi: 10.1109/ICIINFS.2016.8263046.

[39]. S. Garg, S. Urooj, and R. Vijay, "Detection of cervical cancer by using thresholding watershed segmentation," in 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2015, pp. 555–559.

[40]. S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "Fraunhofer Institute for Telecommunications , Heinrich Hertz Institute , Berlin , Germany Department of Electrical Engineering , Technical University of Berlin , Germany .," pp. 1–5.

[41]. "Find Open Datasets and Machine Learning Projects | Kaggle." https://www.kaggle.com/datasets

(accessed Aug. 14, 2020).

[42]. K. Grill-Spector, K. S. Weiner, J. Gomez, A. Stigliani, and V. S. Natu, "The functional neuroanatomy of face perception: from brain measurements to deep neural networks," Interface Focus, vol. 8, no. 4, p. 20180013, Aug. 2018, doi: 10.1098/rsfs.2018.0013.

[43]. "(1) Multimodal Deep Learning for Cervical Dysplasia Diagnosis | Request PDF," ResearchGate. https://www.researchgate.net/publication/308816998_Multimodal_Deep_Learning_for_Cervical_Dysplasia_Diagnosis (accessed Aug. 14, 2020).

[44]. T. Xu et al., "Multi-feature based Benchmark for Cervical Dysplasia Classification Evaluation," Pattern Recognit., vol. 63, pp. 468–475, Mar. 2017, doi: 10.1016/j.patcog.2016.09.027.

[45]. "Bergstra and Bengio - Random Search for Hyper-Parameter Optimization.pdf." Accessed: Aug. 14, 2020. [Online]. Available: https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf.

[46]. M. Z. Alom, M. Hasan, C. Yakopcic, and T. M. Taha, "Inception Recurrent Convolutional Neural Network for Object Recognition," ArXiv170407709 Cs, Apr. 2017, Accessed: Aug. 14, 2020. [Online]. Available: http://arxiv.org/abs/1704.07709.

[47]. G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-Deep Neural Networks without Residuals," ArXiv160507648 Cs, May 2017, Accessed: Aug. 14, 2020. [Online]. Available: http://arxiv.org/abs/1605.07648.

[48]. "(1) (PDF) Neural Network Based Rhetorical Status Classification for Japanese Judgment Documents," ResearchGate. https://www.researchgate.net/publication/338402067_Neural_Network_Based_Rhetorical_Status_Classification_for_Japanese_Judgment_Documents (accessed Aug. 14, 2020).

[49]. "(1) (PDF) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review." https://www.researchgate.net/publication/317496930_Deep_Convolutional_Neural_Networks_for_Image_Classification_A_Comprehensive_Review (accessed Aug. 14, 2020).

[50]. J. Liu, Y. Yang, S. Lv, J. Wang, and H. Chen, "Attention-based BiGRU-CNN for Chinese question classification," J. Ambient Intell. Humaniz. Comput., Jun. 2019, doi: 10.1007/s12652-019-01344-9.

[51]. "(1) (PDF) Performance Analysis of Deep Learning CNN Models for Disease Detection in Plants using Image Segmentation." https://www.researchgate.net/publication/337355921_Performance_Analysis_of_Deep_Learning_CNN_Models_for_Disease_Detection_in_Plants_using_Image_Segmentation (accessed Aug. 14, 2020).

[52]. J. Görtler, R. Kehlbeck, and O. Deussen, "A Visual Exploration of Gaussian Processes," Distill, vol. 4, no. 4, p. e17, Apr. 2019, doi: 10.23915/distill.00017.

[53]. L. Rampasek and A. Goldenberg, "TensorFlow: Biology's Gateway to Deep Learning?," Cell Syst., vol. 2, no. 1, pp. 12–14, Jan. 2016, doi: 10.1016/j.cels.2016.01.009.

[54]. X. Zhang and S.-G. Zhao, "Cervical image classification based on image segmentation preprocessing and a CapsNet network model," Int. J. Imaging Syst. Technol., vol. 29, no. 1, pp. 19–28, Mar. 2019, doi: 10.1002/ima.22291.

[55]. "(1) (PDF) Application of deep learning to the classification of images from colposcopy," ResearchGate.

https://www.researchgate.net/publication/322368324_Application_of_deep_learning_to_the_classificat ion_of_images_from_colposcopy (accessed Aug. 14, 2020).

[56]. "Feature selection considering two types of feature relevancy and feature interdependency | Expert Systems with Applications: An International Journal." https://dl.acm.org/doi/10.1016/j.eswa.2017.10.016 (accessed Aug. 14, 2020).

[57]. "[1411.6228] From Image-level to Pixel-level Labeling with Convolutional Networks." https://arxiv.org/abs/1411.6228 (accessed Aug. 14, 2020).

[58]. X. Q. Zhang and S. G. Zhao, "Cervical image classification based on image segmentation preprocessing and a CapsNet network model," Int. J. Imaging Syst. Technol., vol. 29, no. 1, pp. 19–28, 2019, doi: 10.1002/ima.22291.