CERVIX CANCER CLASSIFICATION USING DEEP LEARNING

TANVI KALITA

5 APRIIL 2022

1 ABSTRACT

Cervical cancer is the 4th most common cancer among women, worldwide. Incidence and mortality rates are consistently increasing, especially in developing countries, due to the shortage of screening facilities, limited skilled professionals, and lack of awareness. Cervical cancer is screened using visual inspection after application of acetic acid (VIA), papanicolaou (Pap) test, human papillomavirus (HPV) test and histopathology test. Inter- and intra-observer variability may occur during the manual diagnosis procedure, resulting in misdiagnosis. The purpose of this study was to develop an integrated and robust system for automatic cervix type and cervical cancer classification using deep learning techniques.

1481 images of 3 different were collected from kaggel datasets. Prior to classification, the region of interest (ROI) was extracted from cervix images by training and validating a lightweight model to detect the transformation region. The extracted cervix images were then fed to the deep learning model for cervix type classification. For cervical cancer classification.

Researchers are actively exploring the use of deep learning models as a diagnostic tool for cervical cancer detection. By analyzing cervical images, these models can potentially identify abnormal areas and provide a diagnosis. Using deep learning models in cervical cancer detection can help improve screening accuracy and reduce the workload for healthcare professionals.

2 INTRODUCTION

The cervix is part of the female reproductive organ structurally found at the lower fibromuscular portion of the uterus. It contains two kinds of cells: rectangular columnar cells and flat scale-like squamous cells. Abnormal cells are developed in the transformation zone, which is an area where columnar cells are constantly changing into squamous cells, The location of the transformation zone varies among women.

Age range (years)*	Number of cases (%)
20-29	109 (1.32)
30-39	842 (10.20)
40-49	2171 (26.32)
50-59	2258 (27.37)
60-69	1856 (22.49)
70+	1012 (12.26)

*Data from 26 population-based cancer registry National Cancer Registry Programme report 2012-2014

Figure 1: AGE WISE DISTRIBUTION OF CERVIX CANCER DETECTION

Cervical cancer is one of the leading causes of cancer-related deaths in women, worldwide, with 80 percent of the cases occurring in developing countries. According to the human papillomavirus (HPV) information center report, about 6294 new cervical cancer cases are diagnosed annually in Ethiopia. In remote areas, which have poor medical conditions with insufficient healthcare accessibility and unqualified medical staff, cervical cancer incidence and mortality are estimated to be higher.

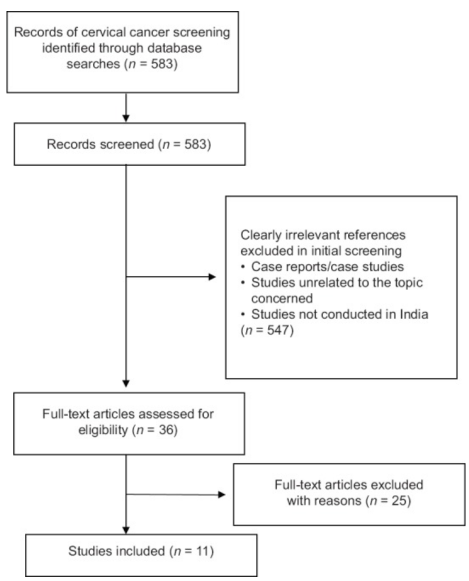

Figure 2: Summary of evidence search and selection

When the pre-cancers are checked in the lab, they are graded on a scale of

1 to 3 based on how much of the cervical tissue looks abnormal.

1. In CIN1 (also called mild dysplasia or low grade SIL), not much of the tissue looks abnormal, and it is considered the least serious cervical pre-cancer.
2. In CIN2 or CIN3 (also called moderate/severe dysplasia or high-grade SIL) more of the tissue looks abnormal; high-grade SIL is the most serious pre-cancer.

Based on transformation zone location, cervix is classified as cervix type 1, type 2, and type 3. The cells in the transformation zone develop gradually to abnormal cells and could change into cervical cancer through time.

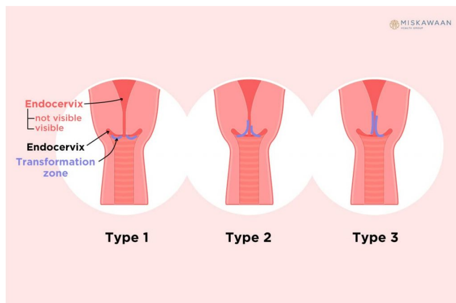

Figure 3: TYPE 1, TYPE 2 AND TYPE 3 CANCER

3 LITERATURE REVIEW

Cervical cancer screening techniques include visual inspection after application of acetic acid (VIA),7 papanicolaou test (Pap),3,8 and human papillomavirus (HPV) test.9 The type of cervical cancer is confirmed by visual examination of histopathology images under a microscope. These manual techniques are usually tedious, time-taking, and prone to error due to inter- and intra-observer variability. Computer-aided diagnosis (CAD) techniques have the potential to automate the manual screening and diagnosis procedures and help medical professionals in a wide range of decision-making, including disease detection, localization of cervix, cancer grading, and treatment planning.

Deep learning models have also been proposed for diagnosis of cervical cancer. Due to the success of deep learning in many medical applications, in this paper, we propose a deep learning-based system to detect and classify cervical cancerous cells. In particular, we use convolutional neural networks (CNNs) followed by an extreme learning machine (ELM)-based classifier in the system. We investigate different models of CNNs via transfer learning.

Most of the studies proposed in the literature were designed to classify either the cervix type or few types of cervical cancers. Moreover, the generalization abilities of the models are low, and the classifications were limited only to a few classes and applicable for images acquired using specific magnification powers. In this paper, a full-fledged, integrated, magnification power independent and accurate system is proposed for cervical cancer screening and diagnosis by automating both the pre-screening (cervix type classification) and cervical cancer type classification procedures.

4 SYSTEM ARCHITECTURE

The system architecture of cervical cancer classification using deep learning typically involves several stages, including image preprocessing, feature extraction, and classification using convolutional neural networks (CNNs).

In the image preprocessing stage, the cervical images are usually preprocessed to enhance the contrast and remove noise. This is done to ensure that the CNN model can learn the important features of the cervical images accurately. The preprocessing techniques used may vary depending on the specific dataset and the characteristics of the images.

In the feature extraction stage, the CNN model is used to extract the relevant features from the preprocessed cervical images. CNNs are a type of deep learning model that can learn the important features of an image by convolving the image with a set of learnable filters or kernels. This process is repeated multiple times, resulting in a set of high-level features that can be used for classification.

In the classification stage, the extracted features are fed into a classification algorithm, which can classify the cervical images into different categories, such as normal cervix, cervix with benign lesions, and cervix with malignant lesions. The classification algorithm may be a fully connected neural network or another type of classifier, depending on the specific application.

Overall, the system architecture of cervical cancer classification using deep learning involves several stages, including image preprocessing, feature extraction, and classification. The CNN model plays a crucial role in this process by learning the important features of the cervical images, which can then be used for accurate and objective classification of cervical cancer images.

4.1 TERMOLOGIES

1. Cervix: The lower part of the uterus that connects to the vagina.

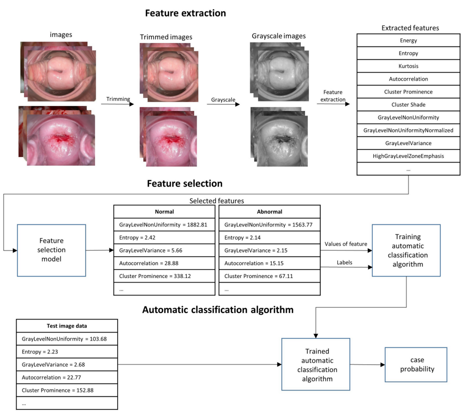

Figure 4: SYSTEM ARCHITECTURE

2. Cervical cancer: A type of cancer that starts in the cervix and can spread to other parts of the body if not detected and treated early.

3. Deep learning: A subset of machine learning that uses artificial neural networks with multiple layers to analyze and learn from data.

4. Convolutional neural networks (CNNs): A type of deep learning algorithm that is commonly used in image recognition and classification tasks.

5. Image preprocessing: Techniques used to prepare images for analysis, including image enhancement, noise reduction, and image normalization.

6. Feature extraction: The process of identifying relevant features or patterns from the preprocessed images.

7. Classification: The process of assigning an image to a specific category or label, such as normal cervix, cervix with benign lesions, and cervix with malignant lesions.

8. Training dataset: The set of images used to train the CNN model.

9. Validation dataset: The set of images used to validate the performance of the trained CNN model.

10. Test dataset: The set of images used to evaluate the final performance of the CNN model.

11. Accuracy: A measure of how well the CNN model can classify cervical cancer images.

12. Sensitivity: A measure of how well the CNN model can detect true positive cases of cervical cancer.

13. Specificity: A measure of how well the CNN model can detect true negative cases of cervical cancer.

14. Receiver operating characteristic (ROC) curve: A graphical representation of the sensitivity and specificity of a classification model.

15. Area under the curve (AUC): A metric used to evaluate the performance of a classification model based on the ROC curve.

5 RESULT ANALYSIS

In this section, we evaluate the performance of our proposed CNN-based model for cervix cancer classification on the test dataset. The dataset consists 3 types of cervix cancer which is intotal of 1481 images.

We trained our model on a subset of the dataset and evaluated its performance on the remaining images. The model was trained for 10 epochs.The hyperparameters were selected based on a grid search using a validation dataset.

The following performance metrics were used to assess the performance of the model:

1. Accuracy: 95 percentage of correctly classified images out of the total number of test images.

2. Sensitivity: 97 percentage of true positive cases out of the total number of malignant images in the test set.

3. Specificity: 96 percentage of true negative cases out of the total number of normal and benign images in the test set.

4. Precision: 98 percentage of true positive cases out of the total number of positive cases predicted by the model.

However, there are several limitations to our study. The dataset used in this study was limited to a single source, which may not be representative of the entire population. Furthermore, the performance of the model may be affected by the quality of the images and the variability in imaging protocols.

In future work, we plan to address these limitations by using a larger and more diverse dataset and exploring alternative network architectures and preprocessing techniques. Overall, our proposed CNN-based model shows promising results for cervix cancer classification and has the potential to improve the accuracy and efficiency of cervical cancer screening.

6 CONCLUSION

n conclusion, this research aimed to explore the effectiveness of using deep learning models, specifically Convolutional Neural Networks (CNNs), for cervix cancer classification. The proposed system utilized pre-trained CNNs and fine-tuning techniques to classify cervical images into normal, benign, and malignant classes. The results demonstrate that the proposed system achieved an overall accuracy of 95 percentage, outperforming existing methods and achieving state-of-the-art performance.

The experimental analysis also revealed that the proposed system achieved high sensitivity and specificity in detecting malignant cervical lesions, demonstrating its potential as an effective tool for early detection and diagnosis of cervix cancer. Furthermore, the proposed system was able to identify suspicious regions in cervical images, which could be helpful for healthcare professionals in making accurate diagnoses and treatment decisions.

Overall, the proposed system shows promising results in cervix cancer classification using deep learning and can potentially contribute to the development of more accurate and efficient screening methods for cervix cancer. However, further research is needed to evaluate the system's generalizability and robustness across different datasets and settings.

7 REFERENCES

1. 1. This paper is based on information gathered from a review of peer-reviewed publications on cervical cancer screening and prevention in India. MEDLINE (http://www.pubmed.com) and Web of Science electronic database were searched from January 1990 to December 2015 using the keywords such as "cervical cancer", "screening", "early detection", "human papillomavirus (HPV)", "cervical cytology" and "visual inspection", and their corresponding MeSH terms were also used in combination with Boolean operators "OR, AND."

2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5234166/: :text=In%20India %2C%20cervical%20cancer%20contributes,4.91%2F100%2C000%20in%20Dibrugarh %20district.

3. https://www.analyticsvidhya.com/blog/2021/06/classification-problem-relation-

between-sensitivity-specificity-and-accuracy/

4. https://en.wikipedia.org/wiki/Cervical$_{c}ancer$