applied sciences

MDPI

Article

Earth Mover's Distance-Based Tool for Rapid Screening of Cervical Cancer Using Cervigrams

Kumar Dron Shrivastav [1], Priyadarshini Arambam [2], Shelly Batra [2], Vandana Bhatia [3], Harpreet Singh [4], Vinita Kumar Jaggi [5], Priya Ranjan [6], Eyad H. Abed [7,*] and Rajiv Janardhanan [8,*]

[1] Laboratory of Health Data Analytics and Visualization Environment, Amity Institute of Public Health, Amity University Uttar Pradesh, Noida 201313, India; kdshrivastav@amity.edu
[2] Batra Hospital and Medical Research Centre, New Delhi 110062, India; priyadarshini.arambam@gmail.com (P.A.); shells1857@gmail.com (S.B.)
[3] Netaji Subhas University of Technology, Delhi 110078, India; vandana.bhatia@nsut.ac.in
[4] Indian Council of Medical Research, New Delhi 110029, India; hsingh@bmi.icmr.org.in
[5] Delhi State Cancer Institute (East), New Delhi 110095, India; drvinitakumar@yahoo.co.in
[6] Bhubaneswar Institute of Technology, Bhubaneswar 752054, India; pranjan@gmail.com
[7] Institute for Systems Research, University of Maryland, College Park, MD 20741, USA
[8] Faculty of Medical & Health Sciences, SRM Institute of Science & Technology, Chennai 603203, India
* Correspondence: abed@umd.edu (E.H.A.); rajivj@srmist.edu.in (R.J.); Tel.: +91-96-5031-9728 (R.J.)

Abstract: Cervical cancer is a major public health challenge that can be cured with early diagnosis and timely treatment. This challenge formed the rationale behind our design and development of an intelligent and robust image analysis and diagnostic tool/scale, namely "OM—The OncoMeter", for which we used R (version-3.6.3) and Linux (Ubuntu-20.04) to tag and triage patients in order of their disease severity. The socio-demographic profiles and cervigrams of 398 patients evaluated at OPDs of Batra Hospital & Medical Research Centre, New Delhi, India, and Delhi State Cancer Institute (East), New Delhi, India, were acquired during the course of this study. Tested on 398 India-specific women's cervigrams, the scale yielded significant achievements, with 80.15% accuracy, a sensitivity of 84.79%, and a specificity of 66.66%. The statistical analysis of sociodemographic profiles showed significant associations of age, education, annual income, occupation, and menstrual health with the health of the cervix, where a *p*-value less than (<) 0.05 was considered statistically significant. The deployment of cervical cancer screening tools such as "OM—The OncoMeter" in live clinical settings of resource-limited healthcare infrastructure will facilitate early diagnosis in a non-invasive manner, leading to a timely clinical intervention for infected patients upon detection even during primary healthcare (PHC).

Keywords: cervical cancer; cervigrams; colposcopy; early detection; screening

check for updates

Citation: Shrivastav, K.D.; Arambam, P.; Batra, S.; Bhatia, V.; Singh, H.; Jaggi, V.K.; Ranjan, P.; Abed, E.H.; Janardhanan, R. Earth Mover's Distance-Based Tool for Rapid Screening of Cervical Cancer Using Cervigrams. *Appl. Sci.* **2022**, *12*, 4661. https://doi.org/10.3390/app12094661

Academic Editors: Keun Ho Ryu and Nipon Theera-Umpon

Received: 23 March 2022
Accepted: 2 May 2022
Published: 6 May 2022

1. Introduction

Cervical cancer is a major unmet clinical need and is a serious public health problem. Early detection is not widely available in developing countries, and the standard detection method (Pap test) is not sufficiently sensitive or specific [1]. According to the World Health Organization (WHO) estimates, cervical cancer is among the highest morbidity- and mortality-causing cancers, with 570,000 new cases annually. Reports by the WHO showed that 77,348 of women affected by cervical cancer died in the recent past, making it the second-highest mortality-causing cancer in India [2,3].

With a diverse genetic base and a resource-limited healthcare infrastructure, India has a disproportionate burden and challenging disease pattern of cervical cancer, necessitating the need to develop robust diagnostic tools to facilitate large-scale screening in order to alleviate the clinical prognosis. The conventional approach for cervical cancer diagnosis is time-consuming and expert-oriented and requires a specific set of resources, leading to misdiagnoses and missed diagnoses, thus resulting in frequent false negatives and false positives [4].

Approximately 453 million women of Indian origin above the age of 15 years are at the risk of being afflicted by cervical cancer. Five percent of these Indian women are estimated to be infected by a human papilloma virus (HPV) serotype (HPV-16/18) at any given instant, and an astounding 83.2% of invasive cervical carcinoma lesions are known to have either the HPV 16 or 18 serotypes [5].

The mortality rate from cervical cancer could be significantly reduced through large-scale screening at both the regional and global levels. A recent report suggested that unimodal screening of large-scale patients using HPV vaccination is indeed insufficient to manage the burden of cervical cancer in a country such as India endowed with a wide genetic base and diverse geological relief structures [6,7].

Cervical cancer, a major challenge for public health professionals and clinicians, requires a reliable and stable technology to enable large-scale rapid screening to reduce its burden. This formed the premise for the development of an interactive and robust diagnostic tool endowed with attributes to efficiently document and analyze the morphological and clinical parameters to accurately detect and diagnose cervical cancer in resource-limited settings [8]. Computational automation of digital colposcopy for its large-scale screening can significantly enhance accuracy and minimize error rates in diagnosis, enabling quicker and timely intervention strategies [9].

The development of an advanced automated diagnostic tool using image analytics would play a key role in the use of non-invasive/minimally invasive technologies as an adjunct clinical aide for facilitating the rapid and precision-oriented screening of cervical cancer [10,11].

Our Contributions

Om—The OncoMeter is an outcome of this targeted research, which consists of an experiment classifying 1481 labelled cervigrams from the Intel Kaggle MobileODT repository as well as 398 cervigrams of individual women collected from Batra Hospital and Medical Research Centre (BHMRC), New Delhi and Delhi State Cancer Institute (East), New Delhi, India. We believe that not only can our open-source software enable accurate disease labeling but also can facilitate triaging of cervigrams lesions in order of their severity. To the best of our knowledge, this is the first report of an Earth Mover's Distance (EMD)-based scale for ranking lesions in order of severity and can provide the rationale for the deployment of a robust and reliable triage system facilitating the large-scale screening of cervical cancer in LMICs. For the first time, a systematic pattern in oncological cervigrams has been developed. Cervigram similarity, which can be calculated using Earth Mover's Distance in an R environment, has been proposed as a measure of oncological development. This measure has been used to develop the oncometer scale to represent the extent of cervical cancer spread, which can be thought of analogous to a thermometer, which gives the extent of a fever. Based on this similarity measurement, one can triage the different cases of cervical cancer in order of their severity. This is a remarkable outcome that came to us by serendipity. A major question asked by oncologists was how do we handle so many normal or healthy cervigrams. The use of Earth Mover's Distance provides an optimal way to rank even healthy cervigrams, where we can go from most healthy to least healthy. This idea is actually used in this work to perform disease tagging.

2. Materials and Methods

In the present study, the digitized cervigrams as well as socio-demographic data (Table 1) of 398 subjects were acquired from the outpatient department (OPD) of BHMRC and DSCI after obtaining prior institutional ethical clearance and informed patient consent. R (version 2021.09.0 Build 351), an open source software tool equipped with highly advanced image analysis tools and techniques, has been used as the programming language. We used R-studio (Integrated Development Environment for R programming language) to automatically demarcate the morphological features, color intensity, sensitivity, lesion detection, contour formation, enhanced visual inspection, and other clinical parameters, as shown in Figure 1. The whole process from conceptualized to its demonstration has been depicted in Data Flow Diagram (DFD) (Figure 1), with an emphasis on the unhindered acquisition of cervigrams (colposcope-derived images) along with its processing and result generation in almost real time using the optimal amount of computational resources.

Table 1. Socio-demographic features of these subjects evaluated at the OPD at BHMRC and DSCI (East): it depicts the socio-demographic features of the subjects/patients evaluated at the OPD of BHMRC, New Delhi, India and DSCI (East), New Delhi, India. The socio-economic demographic/epidemiological data (Table 1) of 398 subjects, was labeled by an expert gynecologist as normal, cervicitis, precancerous, suspected cancerous, abnormal (vaginitis, Nabothian cyst, polyp, and cervical erosion), and different confirmed carcinoma in the cervix, which were undertaken to develop the novel OM—The OncoMeter scale to rank the lesions in the order of their severity. Out of 398 subjects, 102 subjects were found to be normal; 57 had cervicitis; 167 subjects were found to have abnormal lesions; 3 were precancerous; 25 were suspected of cancer and squamous cell carcinoma each; and 5, 9, 3, and 2 had ca cx I, ca cx II, ca cx III, and ca cx IV, respectively. It was observed that age (p = 0.000), education (p = 0.000), occupation (p = 0.001), income (p = 0.000), and menstrual health (p = 0.000) showed statistically significant associations with the health of the cervix, where a p-value less than (<) 0.05 was considered statistically significant.

Category		**Cases (n = 296)**			**Control (n = 102)**	***p* Value**
		Abnormal Cervix (Cervicitis, Vaginitis, Nabothian Cyst, Cervical Erosion)	**Suspected Cancer, Precancerous**	**Ca Cx (Squamous Cell Carcinoma, Ca cx I, Ca cx II, Ca Cx III, Ca Cx IV)**	**Women with Normal Cervix**	
Age	<20	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0.000
	21–40	145 (36.4%)	8 (2.0%)	8 (2.0%)	58 (14.6%)	
	41–60	76 (19.1%)	15 (3.8%)	23 (5.8%)	36 (9.0%)	
	>60	2 (0.5%)	5 (1.3%)	13 (3.3%)	8 (2.0%)	
Education	Illiterate	26 (6.5%)	11 (2.8%)	23 (5.8%)	14 (3.5%)	0.000
	Literate Primary/high school/senior secondary	141 (35.4%)	12 (3.0%)	18 (4.5%)	59 (14.8%)	
	graduation and above	57 (14.3%)	5 (1.3%)	3 (0.8%)	29 (7.3%)	
Occupation	Housewife	200 (50.3%)	25 (6.3%)	38 (9.5%)	89 (22.4%)	0.001
	Government job	3 (0.8%)	1 (0.3%)	0 (0.0%)	4 (1.0%)	
	Private job	20 (5.0%)	0 (0.0%)	2 (0.5%)	9 (2.3%)	
	Other	1 (0.3%)	2 (0.5%)	4 (1.0%)	0 (0.0%)	
Socio-economic Status	Low income (below 1 lakh)	52 (13.1%)	9 (2.3%)	28 (7.0%)	25 (6.3%)	0.000
	Middle income (1–10 lakhs)	169 (42.5%)	18 (4.5%)	15 (3.8%)	76 (19.1%)	
	High income (above 10 lakhs)	0 (0.0%)	1 (0.3%)	1 (0.3%)	1 (0.3%)	
	Do not know	3 (0.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Menstrual health	Normal	11 (2.8%)	3 (0.8%)	2 (0.5%)	38 (9.5%)	0.000
	Abnormal (white discharge, heavy bleeding, irregular menstruation, heavy pain, itching, foul smelling)	192 (48.2%)	22 (6.1%)	40 (10.1%)	51 (12.8%)	
	Post-menopausal	21 (5.3%)	3 (0.8%)	2 (0.5%)	13 (3.3%)	
Total				398		

Earth Mover's Distance (EMD) was used to detect initial variances between the cervigrams of normal subjects to set a threshold value for the cervigrams of the normal cervigrams. The regions of interest of the cropped regions were color-coded using a green channel to enhance the sensitivity of visual inspection and lesion detection because it is the closest channel to luminance. Subsequently, the values obtained from the comparison of both normal and abnormal cervigrams were plotted on a scale virtually stacking the cervigrams based upon the EMD value. This lead to the creation of "OM—The Oncometer", a scale used to rank the cervigrams in order of the disease severity. This will significantly improve rapid screening and early diagnosis in a non-invasive manner.

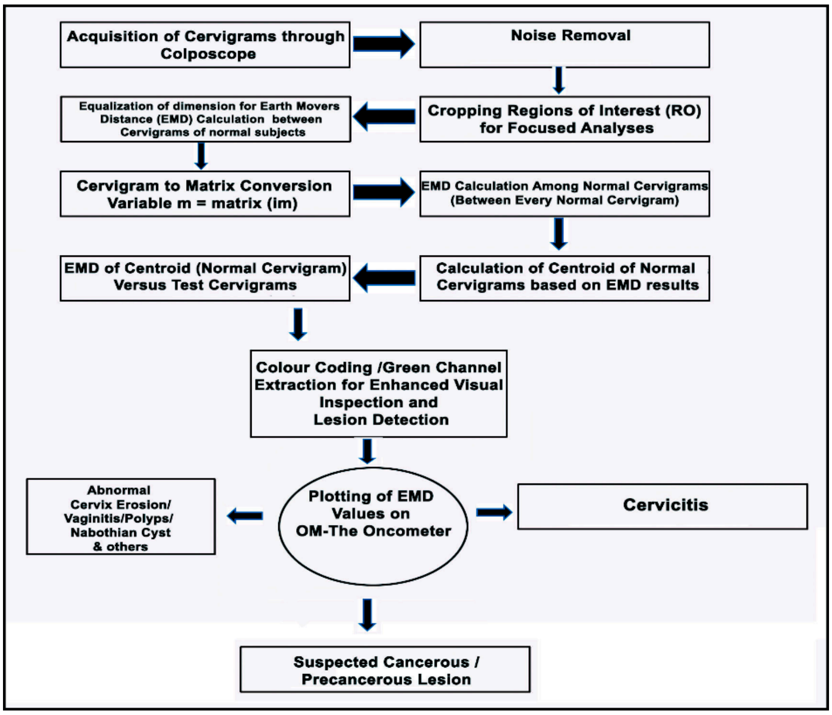


Figure 1. Diagrammatic representation of steps involved in pre-processing and processing of digitized cervigrams for detection of abnormality: our present study depicts a diagrammatic representation of the processing the digitized cervigrams obtained from the patients evaluated at the outpatient department of Batra Hospital and Medical Research Centre. The pre-processing algorithms include noise removal associated with the extraction of the image features (digitized cervigrams in this case) before cropping the raw image of cervigrams to define the regions of interest for equalization of the dimensions before being processed with Earth Mover's Distance.

MySQL Version 14.14 Distrib 5.5.60 has been used for the development of the image/video repository for further processing of digitized cervigrams while LINUX (Ubuntu) version 20.04 has been used as the base computational platform. Digitized Cervigrams acquired using Digital Colposcope (Mobile ODT, Tel Aviv-Yafo, Israel) from anonymized subjects presenting at the OPD of BHMRC and DSCI were subjected to preprocessing algorithms [12] for noise removal before being analyzed for cervical cancer lesion detection using image processing algorithms, as shown in Figure 2a,b.

Apart from a conventional statistical pedagogy, R has numerous dedicated features and capabilities in the area of advanced medical grade image analysis that helped ground the development process of OM—The OncoMeter (Figure 4).

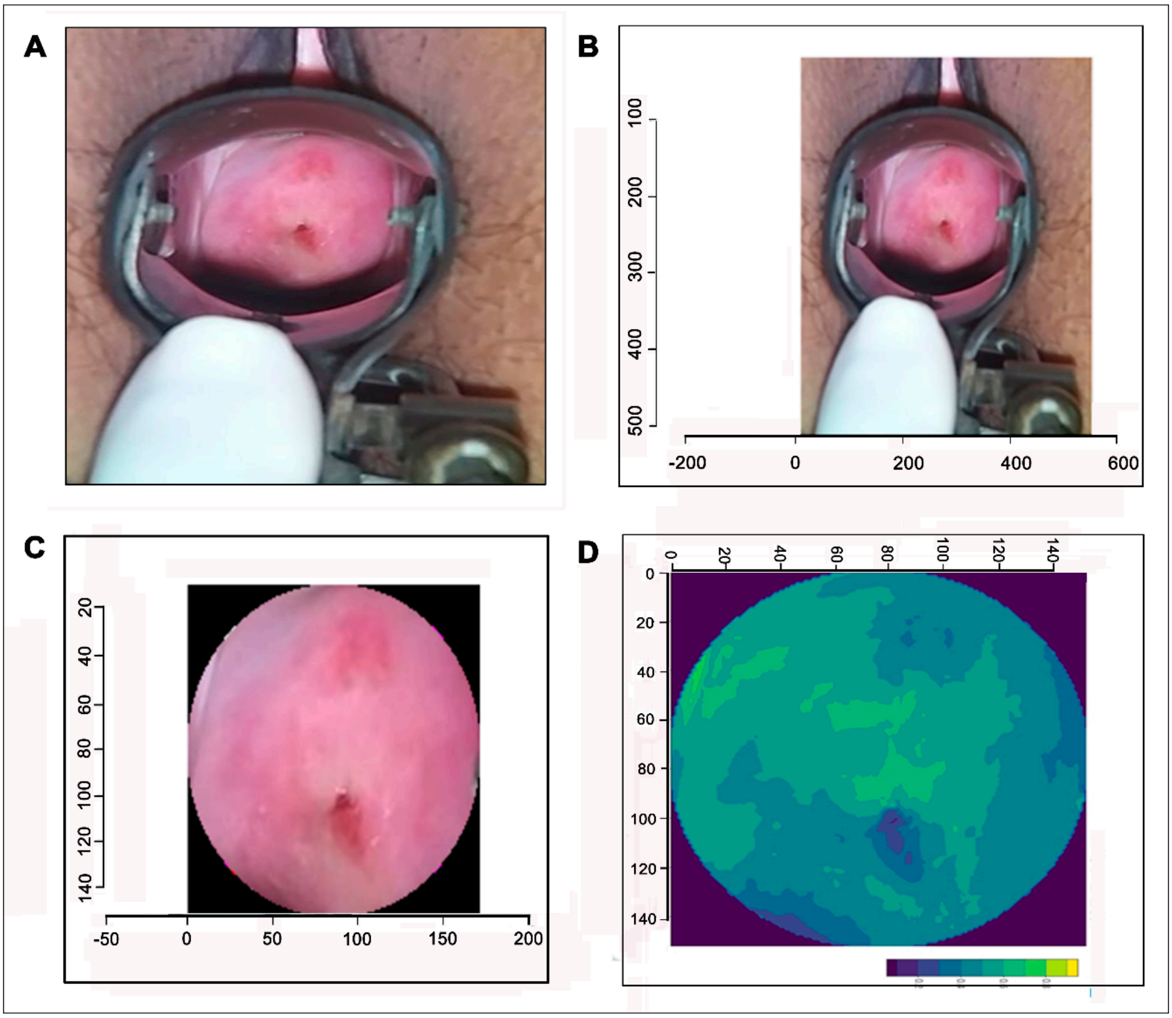
The salient features of the EMD (Earth Mover's Distance) algorithm which allowed for utilization of rigorous quantitative approaches for the colposcopic image comparison and classification are as follows and shown in Figure 4 [13–15]:

- The first and foremost key feature of the algorithm, being free;
- Image similarity-based detection;
- Open source implementation in R;
- Robust application to noisy images;
- Cloud-based storage accessibility and ubiquitous provisioning of services from a collected repository at the local, regional, and global levels;
- This cloud-based information provisioning facilitates automated closed-loop-resource allocation, which in turn provides affordable and accessible healthcare technology platforms for effective management of disease burden.

Training of the algorithm used 81 normal (control) cervigrams of females of Indian origin obtained from the OPD of BHMRC as input (Figure 2). This contributed to the choice of basic threshold parameters in the algorithm. EMD values between every two normal cervigrams from n1 to n81, shown in Figure 3, were calculated and entered into a matrix. EMD was used to calculate the distances between the cervigrams, and based on similarity indices, the cervigrams were categorized to the nearest one. A matrix of 6561 EMD quantitative values was calculated from n1–n2, n1–n3,..., n1–n81, n2–n3, n2–n4, . . . , n2–n81 and continued 95 times until n80–n81 (Figure 3).

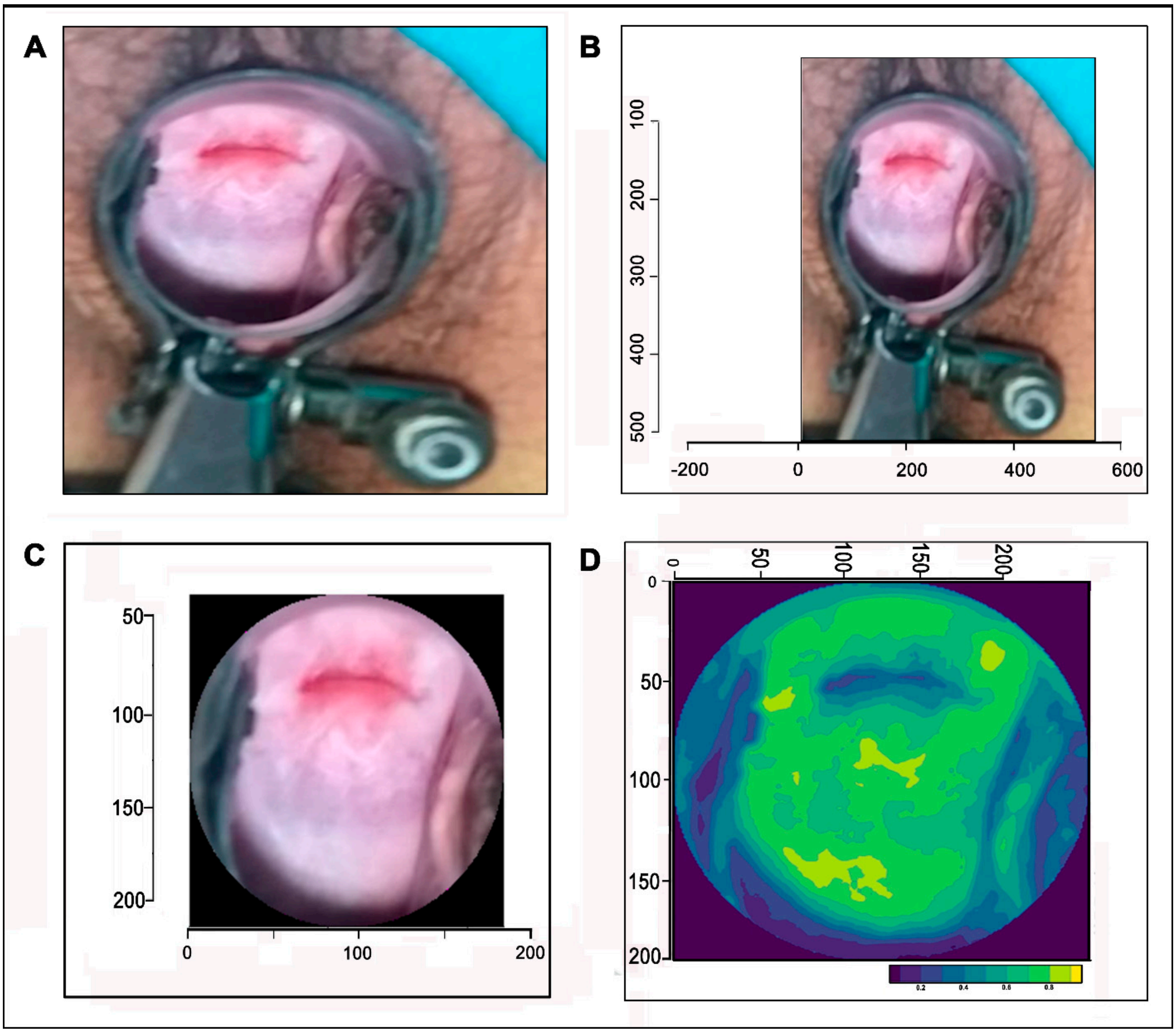
Calculation of the centroid of the normal cervigrams is crucial for designating the threshold values for normal cervigrams. The threshold value of the centroid was taken as a reference value for normal cervigrams for facilitating the EMD-based oncometer scale for ranking the cervical cancer lesions in order of severity (Table 2 and Figure 4).

In addition to the calculation of the centroid for the normal cervigrams, socio-demographic data were analyzed using SPSS version 25.0 (SPSS Inc., Chicago, IL, USA). Descriptive statistics were applied, and bivariate analysis using Chi square test was conducted to detect any significant difference between categorical variables. A p-value less than (<) 0.05 was considered statistically significant. The sensitivity and specificity of the software's capability to detect the cancer cervix lesions were assessed following the protocols of Coughin et al. (1992) [16], and Lalkhen and McCluskey (2008) [17] (Table 3).



(**a**)

Figure 2. *Cont.*



(**b**)

Figure 2. (**a**) **Diagrammatic representation of the image processing of normal cervigrams:** it depicts the processing of the normal cervigrams A by subjecting it to preprocessing algorithms to remove the noise potentially interfering with the extraction of the image features before being processed with green channel color-coding algorithms and the Earth Mover's Distance (EMD) algorithm to facilitate the assignment of a numerical score to the normal cervigrams and to rank them based upon their EMD score. Furthermore, separate images (**A**–**D**) of Figure 2a depicts the raw image (**A**) acquired from colposcope; then the image after preprocessing (**B**) followed by a cropped image, as shown in (**C**); and then the final color-coded image, as shown in (**D**) using green filter in R code, respectively. (**b**) **Diagrammatic representation of the image processing of abnormal cervigrams**: it depicts the processing of the abnormal cervigrams (**A**) by subjecting it to preprocessing algorithms, as shown in (**B**,**C**), to remove the noise potentially interfering with the extraction of the image features before being processed with the green channel, as shown in (**D**), and the color-coding algorithms and Earth Mover's Distance (EMD) algorithm to facilitate the assignation of a numerical score to the abnormal cervigrams and to rank them based upon their EMD score.

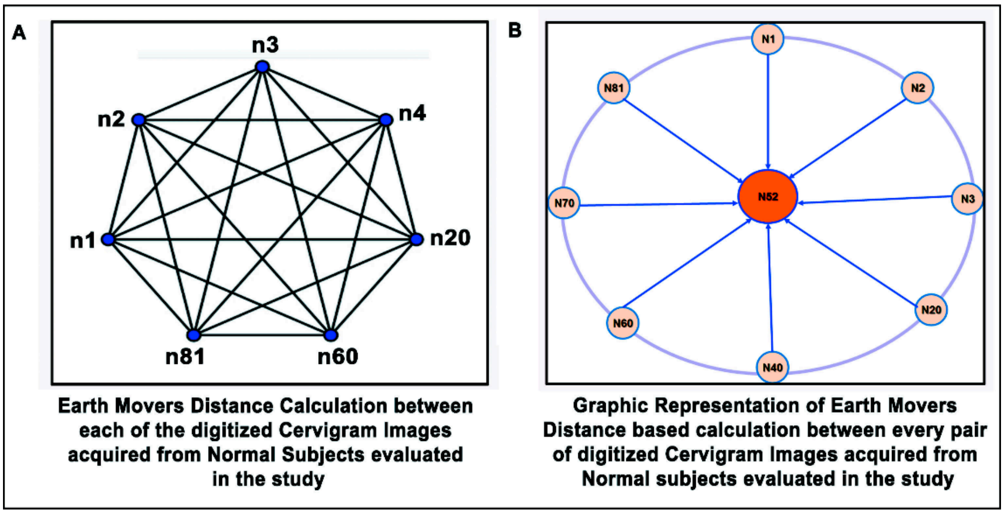


Figure 3. Earth Mover's Distance-based calculation of distances between normal cervigrams among the subjects evaluated at Batra Hospital: in the present study, cervigrams of 81 normal subjects were acquired from the outpatient department of Batra Hospital and Medical Research Centre (BHMRC) after obtaining prior institutional ethical clearance and informed patient consent. Each of the 81 cervigrams obtained from age-matched normal subjects were compared with each other to calculate the variance between the cervigrams of the normal subjects. The Earth Mover's Distance (EMD) enabled calculation of the variance within the cervigrams of the normal subjects and enabled the calculation of the threshold value for the cervigrams of the normal subjects, which in this case, was found to be 26.77 obtained from normal subjects evaluated from the OPD of BHMRC, New Delhi. This threshold value was used to demarcate the normalness of the cervigrams along with its comparison with the abnormal cervigrams as observed in the patients evaluated in the outpatient department of Batra Hospital and Medical Research Centre, Delhi as a part of the current study.

Table 2. Range measurement of different categorical cervigrams: this table depicts the EMD value-based computational tagging of cervigrams into different categories such as normal, abnormal, cervicitis, precancerous, and precancerous/cancerous lesions. The EMD values obtained from the cervigrams form the rationale for developing OM—the OncoMeter, a scale used for computational tagging of the cervigrams in the order of their severity, thereby facilitating the triaging of lesions.

Tagging of Cervigrams	No. of Cervigrams	Accurate Classification of Cervigrams by OM	EMD Range on OM—The OncoMeter
Normal	102	68	0–26.77
Abnormal (Nabothian cyst, vaginitis, cervical erosion, polyp)	167	133	8.162584–146.8711
Cervicitis	57	53	19.41418–140.2637
Precancerous	3	3	34.65366–36.1091
Suspected Cancer	25	21	67.77164–78.53794
Squamous Cell Ca cx	25	24	21.01863–92.59932
Ca cx I	5	3	21.58473–68.71633
Ca cx II	9	9	27.55506–103.9834
Ca cx III	3	3	51.99892–106.1609
Ca cx IV	2	2	42.90317–75.26646
Total	398	319	

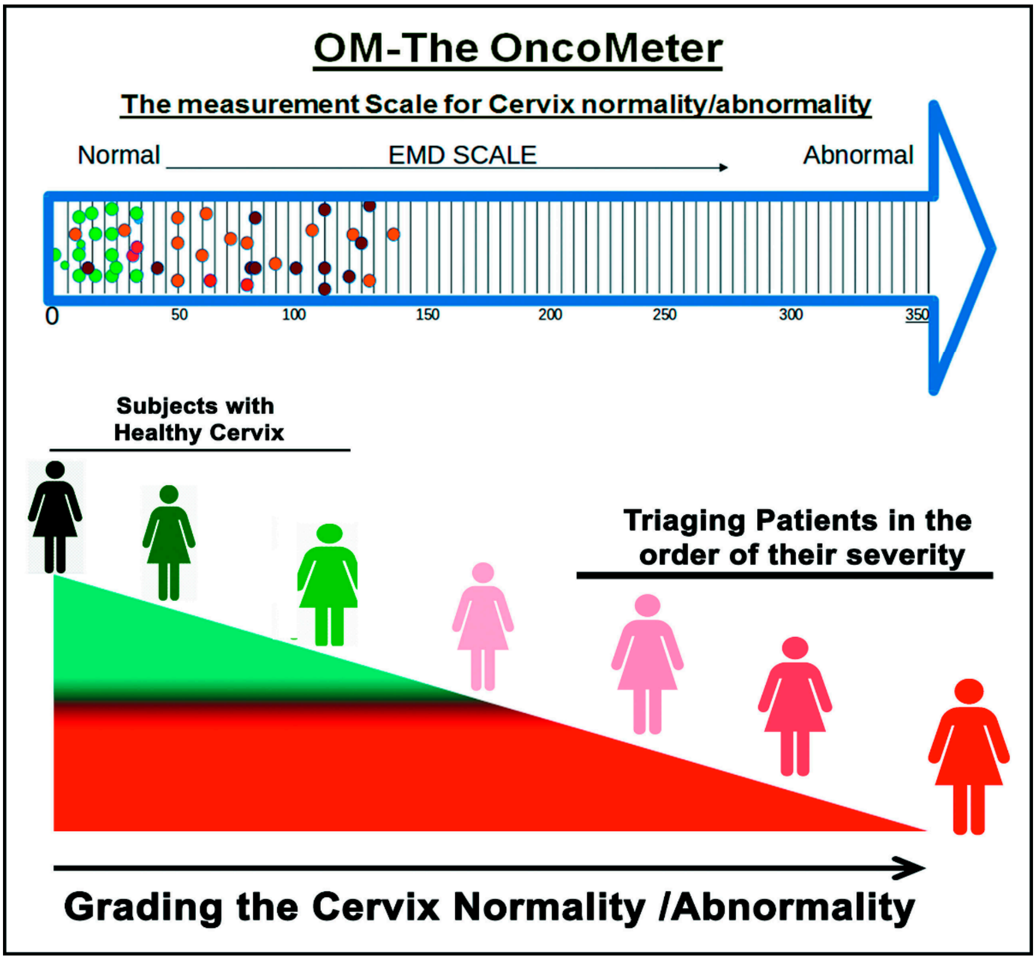


Figure 4. Development of OM—The OncoMeter for triaging of subjects/patients with cervical cancer: this figure summarizes the impact of our study, where we have attempted to make a new scale to virtually rank and grade the cervigrams in the order of their normality or abnormality. This indeed forms the rationale for facilitating the rapid screening of cervical cancer in the rural milieus of the Indian sub-continent besides ranking the cervigram lesions in order of the severity of their disease progression. Such automated processing of chores will ultimately help clinicians to intervene in the patients in need of intervention on a priority basis. The use of this technology will also help in augmenting the hospital-based registry with community-based data, which will indeed provide a more realistic scenario of cervical cancer prevalence in resource-limited healthcare systems prevalent in the Indian sub-continent and elsewhere in the world.

Table 3. Calculation of sensitivity, specificity, and accuracy of the proposed algorithm: we used basic notions of sensitivity and specificity [15,16] to quantify the performance of our proposed algorithm for early detection of cervical cancer as per standard definitions. This table essentially depicts the divergent data included in our study, which was tested by OM—The OncoMeter, and so far, the scale has produced promising results, with 84.79% sensitivity, 66.66% specificity, and 80.15% accuracy.

Cases (Abnormal Cervigrams)—296		Controls (Normal Cervigrams)—102	
True Positive	251	True Negative	68
False Negative	45	False Positive	34
Sensitivity (251/296)	84.79%	Specificity (68/102)	66.66%
Accuracy (319/398)		80.15%	

Calculation of the mean of EMD values was performed in order to obtain the standard normal cervigram for further EMD calculation and for a comparison with other subsets of abnormal cervigrams (Figure 4). Our calculation gave a mean value of 26.77. Cervigram number 52 was designated as the standard cervigram as all the normal cervigrams had the least distances from it with respect to EMD values (Figure 3). The image analytics algorithm-based automated processing of the cervigrams was programmed to obtain simultaneous generations of results so as to provide clinicians with a valuable adjunct clinical decision-making aide.

We used the STARD checklist [18] when writing our report.

3. Results

The socio-demographic data revealed that, among a total of 398 women enrolled for the current study, a vast majority, 219 (55%), were between 20–40 years of age; 74 (18.6%) of the women were illiterate; and 230 (57.7%) of the women were educated up to the high school level. The majority of these women were housewives, 352 (88.5%), with a family annual income between 1–10 lakhs, 278 (69.9%). Only 39 (9.9%) of the women enrolled in the study had attained menopause; 305 (77.2%) of the women presented with the abnormal menstrual conditions (Table 1). A striking observation was that education ($p = 0.000$) and occupation ($p = 0.001$) showed statistically significant associations with the health of the cervix (Table 1).

Furthermore, our analyses indicated that women from lower socio-economic strata belonging to a household with less than INR 100,000 per annum (approximately USD 1300 per annum) had highly unhealthy cervix ($p = 0.000$), indicating poor knowledge of adoptive reproductive and sexual health practices (Table 1), which is indeed responsible for increasing the vulnerability of women to cervical cancer, particularly in LMICs such as India. Age and menstrual health also showed statistically significant association with the health of the cervix.

We believe that the adoption and integration of intelligent decision support systems involving the use of minimally invasive techniques to detect cancerous cervical lesions through automated segmentation of digitized cervigrams on a real time scale would not only form the rationale for the development of effective triage methods towards the early diagnosis of lesions but also prioritize treatment options.

To this end, an EMD-based measurement scale (OM—The OncoMeter) was developed with a threshold value of 26.77 based on cervigrams obtained from healthy subjects evaluated at BHMRC for being designated as normal. Beyond the threshold value of 26.77 (EMD value), the cervigrams were computationally tagged into different categories (Table 2 and Figure 4).

The development of an EMD-based measurement scale for binning and categorization of cervigrams based on their EMD distance from the centroid of normal in a quantitative manner is indeed a stellar achievement in the field of cervigram analysis. The scale was designed for large-scale screening of cervical cancer and to act as an adjunct clinical aide for early and timely diagnosis by clinicians. The measurement scale was illustrated per their categorical EMD values obtained after the designation of the centroid value (26.77), an EMD value based upon the cervigrams obtained from the OPD of BHMRC, as well as its comparison with the abnormal cervigrams (Table 2 and Figure 4).

Sensitivity and Specificity

Current clinical regimens require a large number of clinical tests to confirm or deny the existence of a disease or to refer them for more advanced diagnosis in cases of indecision. The sensitivity of a clinical test is its ability to identify patients afflicted with the disease. The specificity of a clinical test is its ability to identify patients who do not have the disease. Their joint relationship is expressed with RoC (receiver operator characteristic) curves. The sensitivity and specificity of the software's capability to detect the cancer cervix lesions were also assessed following the protocols in [15,16] to quantify the performance of our proposed algorithm for early detection of cervical cancer per standard definitions. Table 3 essentially depicts the divergent data included in our study, which were tested using OM—

The OncoMeter, producing promising results with 84.79% sensitivity, 66.66% specificity, and an accuracy of 80.15%.

Further research with an extensive amount of cervigrams from different regions of the Indian subcontinent encompassing a populace with a divergent genetic base and socio-cultural norms form the rationale for identifying the vulnerable populations to the ravages of this disease. The processing of large numbers of cervigrams would remove the ambiguities/glitches associated with the software.

In other words, further inclusion of data points will make this scale more robust and trustworthy for deployment in the field as an indicator for large-scale disease labeling at the community level.

We believe that with a larger training data-set of higher resolution cervigrams, our algorithms have the potential to perform significantly better.

4. Discussion

An estimated 90% of the globally recorded cervical cancer-related deaths are in low- and middle-income countries (LMICs). Cervical cancer is a public health problem in LMICs such as India, so much so that India alone accounts for one-quarter of the worldwide burden of cervical cancers [19]. It is estimated that cervical cancer will occur in approximately 1 in 53 Indian women during their lifetime compared with 1 in 100 women in more developed regions of the world [20] due to the availability of efficient and accessible screening programs as well as diagnostic and treatment facilities. Apart from the preponderance of HPV infection, a variety of clinical-epidemiological risk factors such as early age of marriage, promiscuous sexual behavior, multiple pregnancies, and poor genital hygiene along with aberrant lifestyle choices such as smoking are known to be associated with increased risk of cervical cancer.

The results obtained from our data indicate that a lack of adequate knowledge about the adoptive reproductive and sexual health practices is indeed responsible for increasing the vulnerability of women to cervical cancer.

Successful implementation of this strategy will have a positive impact on not only understanding the niche-specific drivers for onset and pathological sequelae of cervical cancer but also on providing the rationale for developing novel precision-oriented niche-specific non-linear predictive algorithms for early clinical diagnosis as well as prioritizing the delivery of treatment options to high-risk patients afflicted with cervical cancer. The adoption of low-end mobile health-based applications to propagate Internet and communication technology (ICT)-based awareness about cervical cancer might not only contribute to overcoming region-specific social stigmas and taboos associated with the prevalence of cervical cancer but also facilitate remote connect of cancer afflicted patients with the treating physicians.

The etiopathogenesis of cervical cancer is unique; as a consequence, its characteristics pertaining to different symptoms and parameters necessary for disease labeling, tagging, and further diagnosis are unique and community-specific or niche-specific. One of the biggest challenges is increasing the accuracy of the digitally acquired images, which in turn pertains to low-resolution optics prevailing in colposcopes along with the noise and artifacts such as blood and haziness in the cervigrams (in our case, the MobileODT colposcope had a 13 Megapixel resolution, which needs improvement to visualize cervical cancer lesions more clearly) and to the lack of optimal oncological image/video processing algorithms for outlining a correct set of parameters [21,22]. Some progress has been reported in the histological image of cervical cancer by Taneja et al. [23], where multi-level set-based image processing techniques and deep learning has been used to outline cell nuclei.

Our work should be looked at as complementary or even advanced effort to analyze cervical cancer cervigrams at the normal optical scale, obviating the need for sophisticated microscopy equipment and hence making early cervical cancer detection accessible to economically weaker sections of the society. Modern efforts to leverage the artificial intelligence capabilities in resolving the cervical cancer issue are reported continuously [24].

Disease-labeling of cancer is a data-intensive and rigorous task that requires accurate and specialized expertise.

These skills are not readily available particularly in remote areas with limited healthcare infrastructure. The prevalence of misdiagnoses and missed diagnoses is high in the absence of portable, accurate, and robust diagnostic tools [22], where our proposed tool could aid medical professionals in drastically improving the detection of cervical cancer at an early stage. One of the strategies for ensuring early detection of a cervical cancer lesion pertains to large-scale rapid screening for the disease in PHCs. In the case of cervical cancer, screening is performed through liquid-based cytology and colposcopy. In liquid-based cytology, a major hindrance is the acquisition of a sample and the lack of specialized professionals to make the diagnosis. However, in the recent past, there have been tremendous technological advancements with the advancements of AI (artificial intelligence) [25]. In a resource-limited country such as India, there is an urgent and unmet need to clearly identify reliable and robust visual cues in cervigrams.

Furthermore, these cues should be robust enough to remain invariantly un-occluded under different optical transformations and even in the presence of unwanted objects in images such as hair and other body parts. Such a process will eliminate artifacts and other irrelevant features to seamlessly segment structures for grading measurements of cervical cancer lesions. The automation of this feature ensures that the algorithm detects structural aberrations in cervigrams in a precise and timely manner to make cervical cancer detection even more reliable [26].

The optical clarity of an image is paramount for visual inspection by a camera with direct human eye inspection. Optical aberration, color misrepresentation, and specular reflection in cervigrams are other challenges associated with conventional colposcopes [27]. Our approach takes care of these problems by facilitating the provisioning of much sharper and focused images necessary for the delineation and identification of cancerous lesions. Using image processing-based automated detection of cervical cancer can increase precision and minimize errors. When detected and managed at an early stage, the clinical prognosis for cervical cancer is favorable [28].

This indeed necessitates the need for a technology-driven approach for developing an effective and precision-oriented triage system for facilitating not only large-scale screening but also prioritizing patients for therapeutic intervention. This can minimize the recovery period through screening and follow-up. Taking into account the conditions, parameters, limitations, and the knowledge sharing from manual to computational colposcopy, we tried to incorporate ubiquitous information related to the labeling/tagging of the cervix and have developed an artificial intelligence-enabled intelligent decision support system capable of serving as an adjunct clinical aide for facilitating rapid screening for the detection of cancerous lesions. In this study, we introduced an algorithm-based colposcopic image (cervigrams) analysis and diagnostic technology that uses the attribution method to identify the cervix types of Indian subjects [29,30].

The R software (version 2021.09.0 Build 351) libraries included advanced tools for medical image analysis, which deal with the quantification of the colposcope-derived cervigrams, thereby reducing the burden and time of medical professionals by representing images as numerical data for analysis apart from conventional image processing procedures [31,32].

Limitations:

- The computational part of the proposed algorithm heavily relies on cervigrams acquired by a colposcope; hence, its ability to correctly tag cervical conditions is limited by the truthfulness and resolution of the cervigrams.
- This tool was built on the basis of training data, which were used for concept development and its validated implementation. Large numbers of heterogeneous cervigrams from all across the country are required for refining the tool and then should be deployed in the community with proper settings for other lifestyle and genetic parameters.

5. Conclusions

In this study, multi-modal multi-sensor fusion technology was developed by integrating signals from artificial intelligence-enabled image analyses, HPV serotyping, as well as liquid-based cytology to ensure precision-oriented large-scale rapid screening of subjects at the community level and to eliminate missed diagnoses and misdiagnoses of cervical cancer (Figure 5).

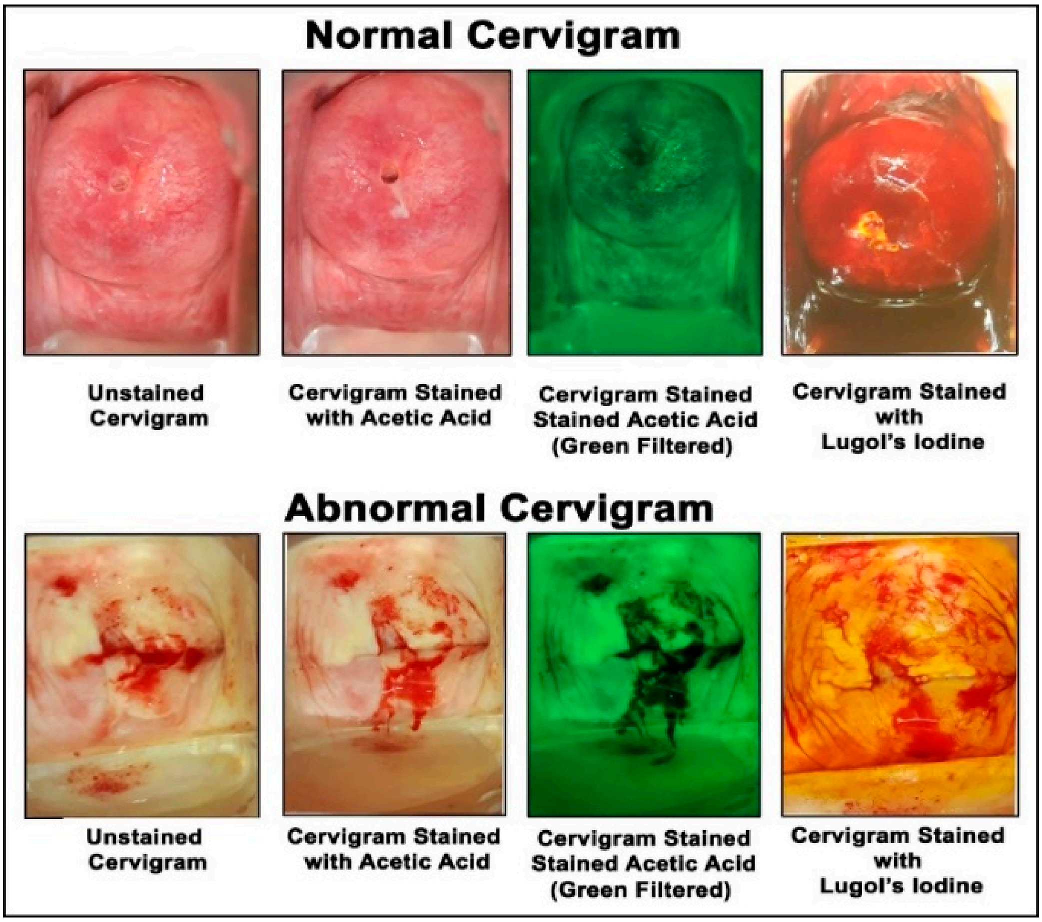


Figure 5. It shows the cervigrams stained with acetic acid (VIA) and Lugol's iodine VILI to ascertain that the lesions stained with VIA are indeed cancerous, as exemplified by dysplasia Lugol's iodine stain. The Lugol's iodine stain works on the principle that the normal epithelial cells contain glycogen while the dysplastic and invasive cancer cells contain little or no glycogen. This is thought to be due to the Warburg effect of increased cytosol glycolysis consequent to the genomic chaos seen in the cancer cell. Iodine is glycophilic and forms tri-iodide molecules within the glycogen polymer spiral. This results in mahogany brown staining of normal epithelial cells. Areas of dysplasia and invasive cancer do not take up iodine as they lack glycogen and appear as pale mustard-colored areas. Modules of health literacy about adoptive sexual and reproductive health practices will also be integrated into the software solution to alleviate the burden of this disease in vulnerable populations belonging to the lowest socioeconomic strata.

The core idea behind this multimedia tool was to help women with lower incomes and more health needs from LMICs, but our vision was to implement it also in developed countries. It primarily targets patients as a self-diagnosing device, and clinicians can use this system as an aide in their assessments.

6. Patents

Copyright was awarded for "Computational Detection and Ranking of Cervical Cancer from Cervigrams" (copyright number: L-78200/2018; diary number 11446/2018-CO/L).

Author Contributions: Conceptualization, K.D.S., P.R. and R.J.; methodology, K.D.S., P.A.,V.K.J., V.B., H.S., S.B., V.K.J., P.R., E.H.A. and R.J.; software, K.D.S., P.R.; validation, S.B. and V.K.J.; formal analysis, K.D.S., P.R. and R.J.; investigation, K.D.S., P.A., P.R. and R.J.; resources, K.D.S., S.B., P.A., V.K.J., H.S., V.B., P.R. and R.J.; data curation, K.D.S., P.R., H.S., E.H.A. and R.J.; writing—original draft preparation, K.D.S., P.A., P.R. and R.J.; writing—K.D.S., P.R. and R.J.; visualization, K.D.S., H.S., P.R. and R.J.; supervision, K.D.S., H.S., E.H.A., P.R. and R.J.; project administration, P.R., S.B., V.K.J., H.S. and R.J.; funding acquisition, K.D.S., P.R. and R.J. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge the Indian Council of Medical Research, New Delhi for support as a grant in aid to Rajiv Janardhanan and Priya Ranjan (grant id No.-2029-0416-No.- ISRM/12(23)/2019) and ICMR Senior Research Fellowship (ref. No- BIC/11(06/2016)) to Kumar Dron Shrivastav.

Institutional Review Board Statement: Ethics approval was received from the ethics committees of Amity University; Noida Uttar Pradesh; Batra Hospital; Medical Research Centre, New Delhi; and Delhi State Cancer Institute (East), New Delhi.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data-sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: We acknowledge the Indian Council of Medical Research, New Delhi for support as a grant in aid to Rajiv Janardhanan and Priya Ranjan (grant id no.-2029-0416-No.- ISRM/12(23)/2019). ICMR Senior Research Fellowship (ref. no- BIC/11(06/2016)) awarded to Kumar Dron Shrivastav and the support of mobile ODT for the provision of cervigrams obtained through the mobile ODT EVA system at BMHRC and DSCI (East) under the supervision of Shelly Batra and Vinita Kumar Jaggi.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rao, J.; Escobar-Hoyos, L.; Shroyer, K.R. Unmet clinical needs in cervical cancer screening. *MLO Med. Lab. Obs.* **2016**, *48*, 8–10. Available online: https://www.mlo-online.com/ (accessed on 22 March 2022). [PubMed]
2. World Health Organization. Who Cervical Cancer. Early Diagnosis and Screening. Available online: www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/ (accessed on 22 March 2022).
3. The Global Cancer Observatory. India Factsheet, WHO. Available online: https://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf (accessed on 22 March 2022).
4. Shrivastav, K.D.; Das, A.M.; Singh, H.; Ranjan, P.; Janardhanan, R. Classification of colposcopic cervigrams using EMD in R. In Proceedings of the International Symposium on Signal Processing and Intelligent Recognition Systems, Bangalore, India, 9–22 September 2018; pp. 298–308. [CrossRef]
5. ICO/IARC. ICO/IARC Information Centre on HPV and Cancer, India Human Papillomavirus and Related Cancers, Fact Sheet. 2018. Available online: https://hpvcentre.net/statistics/reports/IND_FS.pdf (accessed on 22 March 2022).
6. Kumar, P.; Gupta, S.; Das, A.M.; Das, B.C. Towards global elimination of cervical cancer in all groups of women. *Lancet Oncol.* **2019**, *20*, e237. [CrossRef]
7. Shrivastav, K.D.; Taneja, N.; Das, A.M.; Rana, S.; Ranjan, P.; Singh, H.; Jaggi, V.K.; Janardhanan, R. Socio-demographic and Clinico-pathological Profile of Cervical Cancer Patients at a Tertiary Care Centre in New Delhi: A Five-Year Retrospective analysis. *Indian J. Community Health* **2021**, *33*, 634–639. [CrossRef]
8. Tian, R.; Cui, Z.; He, D.; Tian, X.; Gao, Q.; Ma, X.; Yang, J.; Wu, J.; Das, B.C.; Severinov, K.; et al. Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles. *Carcinogenesis* **2019**, *40*, 1220–1228. [CrossRef]
9. Park, S.Y.; Follen, M.; Milbourne, A.; Rhodes, H.; Malpica, A.; MacKinnon, N.B.; MacAulay, C.; Markey, M.K.; Richards-Kortum, R.R. Automated image analysis of digital colposcopy for the detection of cervical neoplasia. *J. Biomed. Opt.* **2008**, *13*, 014029. [CrossRef]
10. Lange, H.; Wolters, R.H. Uterine cervical cancer computer-aided-diagnosis (CAD). *J. Biomed. Opt.* **2010**, 1–10. Available online: https://patents.google.com/patent/US7664300B2/en (accessed on 22 March 2022).

11. de Castro Hillmann, E.; Bacha, O.M.; Roy, M.; Paris, G.; Berbiche, D.; Nizard, V.; Ramos, J.G. Cervical digital photography: An alternative method to colposcopy. *J. Obstet. Gynaecol. Can.* **2019**, *41*, 1099–1107. [CrossRef]
12. Agarwal, A.; Ali, F.; Kopparthi, A.; Ranjan, P.; Shrivastav, K.D.; Janardhanan, R. Specular reflection removal in cervigrams. In *Contemporary Issues in Communication, Cloud and Big Data Analytics*; Springer: Singapore, 2022; pp. 423–437.
13. Rubner, Y.; Tomasi, C.; Guibas, L.J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **2000**, *40*, 99–121. [CrossRef]
14. Boltz, S.; Nielsen, F.; Soatto, S. Earth mover distance on superpixels. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 4597–4600. Available online: https://ieeexplore.ieee.org/abstract/document/5651708/metrics#metrics (accessed on 22 March 2022).
15. Sun, Y.; Lei, M. Method for optical coherence tomography image classification using local features and earth mover's distance. *J. Biomed. Opt.* **2009**, *14*, 054037. [CrossRef]
16. Coughlin, S.S.; Trock, B.; Criqui, M.H.; Pickle, L.W.; Browner, D.; Tefft, M.C. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *J. Clin. Epidemiol.* **1992**, *45*, 1–7. [CrossRef]
17. Lalkhen, A.G.; McCluskey, A. Clinical tests: Sensitivity and specificity. *Contin. Educ. Anaesth. Crit. Care Pain* **2008**, *8*, 221–223. [CrossRef]
18. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; De Vet, H.C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Clin. Chem.* **2015**, *61*, 1446–1452. [CrossRef]
19. Forhan, S.E.; Godfrey, C.C.; Watts, D.H.; Langley, C.L. A systematic review of the effects of visual inspection with acetic acid, cryotherapy, and loop electrosurgical excision procedures for cervical dysplasia in HIV-infected women in low-and middle-income countries. *JAIDS J. Acquir. Immune Defic. Syndr.* **2015**, *68*, S350–S356. [CrossRef]
20. Institute for Health Metrics and Evaluation. The Challenge Ahead: Progress in Breast and Cervical Cancer. *Institute of Health Metrics and Evaluation, IHME.* Available online: https://www.healthdata.org/policy-report/challenge-ahead-progress-and-setbacks-breast-and-cervical-cancer. (accessed on 22 March 2022).
21. Huh, W.K.; Papagiannakis, E.; Gold, M.A. Observed colposcopy practice in US community-based clinics: The retrospective control arm of the IMPROVE-COLPO study. *J. Low. Genit. Tract Dis.* **2019**, *23*, 110. [CrossRef]
22. Hu, L.; Bell, D.; Antani, S.; Xue, Z.; Yu, K.; Horning, M.P.; Gachuhi, N.; Wilson, B.; Jaiswal, M.S.; Befano, B.; et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI J. Natl. Cancer Inst.* **2019**, *111*, 923–932. [CrossRef]
23. Taneja, A.; Ranjan, P.; Ujlayan, A. Multi-cell nuclei segmentation in cervical cancer images by integrated feature vectors. *Multimed. Tools Appl.* **2018**, *77*, 9271–9290. [CrossRef]
24. Menezes, L.J.; Vazquez, L.; Mohan, C.K.; Somboonwit, C. Eliminating cervical cancer: A role for artificial intelligence. In *Global Virology III: Virology in the 21st Century*; Springer: Cham, Switzerland, 2019; pp. 405–422.
25. CDC. Cervical Cancer Screening Guidelines for Average-Risk Women, CDC. Available online: https://www.cdc.gov/cancer/cervical/pdf/guidelines.pdf (accessed on 22 March 2022).
26. Xue, Z.; Antani, S.; Long, L.R.; Jeronimo, J.; Thoma, G.R. Comparative performance analysis of cervix ROI extraction and specular reflection removal algorithms for uterine cervix image analysis. In *Medical Imaging 2007: Image Processing*; International Society for Optics and Photonics: Bellingham, WA, USA, 2007; Volume 6512, p. 65124I.
27. Al-shamasneh, A.R.; Obaidellah, U.H. Artificial intelligence techniques for cancer detection and classification: Review study. *Eur. Sci. J.* **2017**, *13*, 342–370.
28. Dvir, H.; Gordon, S.; Greenspan, H. Illumination correction for content analysis in uterine cervix images. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, 17–22 June 2006; IEEE: Piscataway, NJ, USA, 2006; p. 95.
29. Kim, E.; Huang, X. A data driven approach to cervigram image analysis and classification. In *Color Medical Image Analysis*; Springer: Dordrecht, The Netherlands, 2013; pp. 1–13.
30. Huang, X.; Wang, W.; Xue, Z.; Antani, S.; Long, L.R.; Jeronimo, J. Tissue classification using cluster features for lesion detection in digital cervigrams. In *Medical Imaging 2008: Image Processing*; International Society for Optics and Photonics: Bellingham, WA, USA, 2008; Volume 6914, p. 69141Z.
31. Muschelli, J.; Gherman, A.; Fortin, J.P.; Avants, B.; Whitcher, B.; Clayden, J.D.; Caffo, B.S.; Crainiceanu, C.M. Neuroconductor: An R platform for medical imaging analysis. *Biostatistics* **2019**, *20*, 218–239. [CrossRef] [PubMed]
32. Xu, T.; Zhang, H.; Xin, C.; Kim, E.; Long, L.R.; Xue, Z.; Antani, S.; Huang, X. Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognit.* **2017**, *63*, 468–475. [CrossRef] [PubMed]