Taylor & Francis
Taylor & Francis Group

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/taut20

An automated cervical cancer detection scheme using deeply supervised shuffle attention modified convolutional neural network model

Kanimozhi T. & Vijay Franklin J.

To cite this article: Kanimozhi T. & Vijay Franklin J. (2023) An automated cervical cancer detection scheme using deeply supervised shuffle attention modified convolutional neural network model, Automatika, 64:3, 518-528, DOI: 10.1080/00051144.2023.2196114

To link to this article: https://doi.org/10.1080/00051144.2023.2196114

Published online: 05 Apr 2023.

AUTOMATIKA
2023, VOL. 64, NO. 3, 518–528
https://doi.org/10.1080/00051144.2023.2196114

REGULAR PAPER

OPEN ACCESS Check for updates

An automated cervical cancer detection scheme using deeply supervised shuffle attention modified convolutional neural network model

T. Kanimozhi and J. Vijay Franklin

Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Erode, India

ABSTRACT

Cervical malignant growth is the fourth most typical reason for disease demise in women around the world. In developing countries, women don't approach sufficient screening methods because of the costly procedures to undergo examination regularly, scarce awareness and lack of access to the medical centre. Recently, deep learning-based radiomic methods have been introduced in differentiating vessel invasion from non-vessel invasion in Cervical Cancer (CC) by multi-parametric Magnetic Resonance Imaging (MRI). However, this model doesn't produce sufficient results. In this work, the MRI images are initially pre-processed using bilateral filtering. After pre-processing, the image is segmented by modified U-Net model in order to identify the cancerous region. Extraction of deep semantic information from images by using residual blocks in the processes of contractions and expansions. The last layer of the contracting route uses tightly coupled convolutions in the second phase to speed up feature recycling and feature propagation. It was inferred from the observations that the proposed model was effective as a predictive tool for detecting vessel invasions in preoperative early stages of CC. Proposed model produces 94.00% detection accuracy which is better than the other existing methods.

ARTICLE HISTORY
Received 13 December 2022
Accepted 23 March 2023

KEYWORDS
Cervical cancer; MRI data; pre-processing; modified U-net model; residual blocks and densely connected convolutions

1. Introduction

CCs are an important reason behind cancer deaths in women and developing nations contribute over a quarter of its global incidence [1]. In India, despite the frightening high numbers, government-sponsored screening programmes all over the nation are unheard of. In order to provide proof-based suggestions for using the highly practical screening test to be used in resource-deficient environments, this research was carried out to estimate the incidence of CCs in India and analyse the performance aspects of the existing CC screening instruments. Among all malignancies, the incidence of CCs in Indian women ranges from 6.2% to 62.9% [2].

The greatest age-adjusted incidence rate for CCs is 23.07/100,000 in the state of Mizoram, while the lowest is 4.91/100,000 in the district of Dibrugarh. Visual inspection with acetic acid (VIA), magnified VIA, visual inspection with Lugol's iodine, cytology (Pap smear) and human papillomavirus DNA were all used to predict the sensitivity and specificity of the tests as follows: 67.65% and 84.32%, 65.36% and 85.76%, 78.27% and 87.10%, 62.11% and 93.51%, and 77.81% and 91.54%, correspondingly.

MRI of the pelvis is highly trustworthy imaging modalities for staging, therapy planning and follow-up of CCs. Especially, clinical staging of CCs depends on the nodal status [3] and the volume of tumour. But, this hugely depends on forming accurate segmentation approaches of CCs on MR images, which are efficient in capturing the huge difference in the shape, position and size of the tumour. Certain fully automated segmentation techniques have found their application in the diagnosis of CCs. In recent works used methods like an artificial neural network, Multi-layer Perceptron, Naïve Bayes and decision tree for cervical cancer detection. However, these methods could not learn the multimodal information of CCs which leads to produce more false positive rate.

To overcome this issue in existing work introduced deep learning-based radiomic methods in differentiating vessel invasion from non-vessel invasion in cervical cancer with multi-parametric MRI data. However, this model does not produce sufficient accuracy results. To avoid this problem this work aimed to propose a schema based on radiomic methods using deep learning techniques (DLTs) for identifying vessel invasion and non-vessel invasion in CCs using Magnetic Resonance Imaging (MRI) data using deeply supervised shuffle attention modified convolution neural network (DSSAMCNN).

In this work, the MRI images are initially pre-processed using bilateral filtering to remove any extraneous sounds and other effects. After pre-processing,

CONTACT T. Kanimozhi kani16.cse@gmail.com Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamil Nadu, 638401, India

the picture is segmented with the use of a modified U-Net model to identify the cancerous zone. Extraction of deep semantic information from pictures is done using residual blocks in the processes of contractions and expansions. The last layer of the contracting route uses tightly coupled convolutions in the second phase to speed up feature recycling and feature propagation. In order to lower the likelihood of the vanishing gradient problem and avoid overfitting, three more outputs are employed as deep supervision prior to every higher sampling. It was inferred from the observations that the proposed model was effective as a predictive tool for detecting vessel invasions in the preoperative early stages of CCs.

In this study, automated and enhanced multi-label segmentation of CCs was developed and analysed using models based on DLTs where U-Nets, shuffle attention (SAs) and deeply supervised SAs (DSSAs) were redesigned. U-Net architectures combined benefits of spatial and channel attention methods, combining quantitative, observational, dosimetric and time-based criteria in assessing segmentation efficacies.

The remaining portion of the research is structured as given. Section 1 provides an outline of the available solutions developed for the detection and classification of CCs. Section 2 presents the related works. Section 3 produces the proposed DLTs where detections and segmentations were elaborately discussed. Section 4 studies the performance analysis of the proposed pipeline, and the performance comparison between the proposed modules and benchmark models in the literature is studied. Comparison in terms of computational efficacy, detection and classification accuracy is carried out to focus on the excellence of the proposed schema based on DLTs. Section 5 provides the conclusion of the research where the directions of the future work are provided and the remarks are studied.

2. Literature review

Extremely deep residual learning-based networks were presented by Adweb et al. [4] to enable the screening of CCs and the study investigated performances of ResNets' (residual networks') where different activation functions were generated for the same topology. The proposed residual networks with leaky and parametric rectified linear units (Leaky-RELU and PRELU) activation functions performed in similar fashions with accuracy percentages of 90.2 and 100 during Training/testing on colpos copy cervical image datasets.

Automated segmentations were executed on larger overlaps with ground truth volumes in the study of Sartor et al. [5]. Convolution neural networks (CNNs) can aid with faster and more accurate volume representation. The CNNs are modifications to a typical segmentation network designed to use less memory used CT (Computerized Tomography) structural sets of 191 anorectal tumour cases between 2014 and 2018 that were treated with radiations at Skne University Hospital along with 75 CCs. Structural information of bladders, bowel bags, left/right femoral heads and CTVNs were evaluated clinically for lymph node volumes. The study's automated segmentations were analysed qualitatively by oncologists and assessed in terms of dice score precisions and MSDs (mean surface distances) in mm.

Devi et al. [6] focused on the benefits of Neutrosophic Graph Cut-based Segmentations (NGCSs) on previously processed cervical images. In order to increase classification accuracy, our NGCS-based segmentation largely looks at the overlapping contexts of pre-processed pictures from cervical smears. Segmentations based on NGCSs integrated intensities and spatial information of image pre-process outputs using Neutrosophic sets and in-determinant filter values which were crucial for reducing intensity's indeterminacy values on spatial data. The study's results of CCs detections based on NGCSs when contrasted with traditional graph-cut systems detecting cancers showed improved averages by 13%. Intensity-modulated radiation treatment for CCs toxicity events was reduced as a result of Rigaud et al.'s [7] use of DLTs for automated segmentation to assist in creating regular online dose optimization procedures. In CC patients receiving radiation,

Gou et al. [8] proposed attention-based multi-view feature segmentations called MVFA-Net to overcome the aforementioned issues in the context of 3D information. Features were extracted from weakly correlated neighbouring MRI slices using different volumetric image views, processed individually and finally integrated using channel attentions. The study's MVFA-Net was evaluated on cervical MR data set worth details of 160 CCs patients where their suggested MVFA-Net outperformed other eight medical image segmentation networks, producing better results in terms of DSCs (Dice similarity coefficients) and ASDs (Average surface distances) having values of 2.6–11.1% and 0.39–0.97 mm, respectively. In order to automatically segment cervical MRI volumes.

Luet al. [9] introduced an AugMS-Nets (augmented multiscale networks) that rely on 3D U-Net. A new 3D module is studied for passing through very fine multiscale delineations with a multiscale mechanism in mind as one of the potential methods to handle small item identification. Additionally, a hierarchical deep multiscale supervision method is employed to supervise the side outputs concurrently. The assessment of AugMS-Net on two datasets, one acquired from CT volumes for the liver and the other from MRI volumes for the cervical region, illustrates how generalizable this model is.

Using the INGC-GDES technique, Rajarao and Singh [10] suggested enhanced graph-cut normalizations

for better segmentations and detections of cytoplasm and pap smear cell nucleus borders resulting in accurate recognitions of CCs. The pap smear cervix cells were subjected to the suggested INGC-GDES approach so that its overlapping and hazy borders could be investigated for enhanced identification of CCs Cells. The pre-processed cervical images were transformed into enhanced normalized graph-cut sets and aggregated to utilize benefits of spatial, intensity information of processed images in evaluations. One of the better methods is CC's wireless network-based MRI image segmentation technology. However, information processing and inspection are not capabilities of ordinary technology. Data processing mistakes and other issues might arise if the data are only processed manually. As a result, Liang et al. [11] investigated the CCs' MRI image segmentation technique. They employed DLTs for calculations and analyses which relied on wireless components. The data processing strength and capacity may be improved by utilizing wireless network and computerized algorithm structures.

The effectiveness of U-Net for localizations and segmentations of cervical cancers in MRI images along with the dependability of collected radiomics characteristics like apparent diffusion coefficients (ADCs) were built and evaluated by Lin et al. [12]. The study made use of MR scans from 169 patients with CCs stage IB-IVA; DW (diffusion-weighted) pictures from 144 patients were taken into consideration for training, and a further 25 patients were taken into consideration for testing. To carry out automatic tumour segmentation, a U-Net convolution network was created. For comparison, the manually represented tumour area was used as the baseline. Performance in segmentation was assessed when various input source fusions were used to train. Pearson correlation was used in the extraction and assessment of ADC radiomics. The duplicability of the training was also studied.

For cervical cell pictures, Lakshmi and Ravi [13] devised a segmentation technique. Segmenting and pre-processing cervical images are two procedures that this research takes to accomplish its goal. Utilizing adaptive median filters, cervical image pre-processes attempt to remove noises. The cervical pictures' uneven staining impact is addressed with Haar wavelets. The bit slice plane method is used to achieve the background subtraction. Intersecting Cortical Model (ICM) and cuckoo searches are used to segment the pre-processed pictures, with the latter setting the parameters of ICM to fully automate the procedure.

A cervical segmentation approach that relies on the HSV colour mode was examined by Bai et al. [14] and can aid in the partitioning and extraction of the cervical area from a medical and anatomical standpoint. First, the colposcopic image's histogram (Y) is analysed using the histogram threshold approach. The pre-processed RGB photos are worked on in the second stage. Colposcopic images were converted into their corresponding HSV colour spaces and V component features were extracted using K-means algorithm. Subsequently, segmented cervical regions were obtained using area filters that smoothened edges. The study's experimental results for accuracy, specificity and sensitivity were 87.25%, 81.99% and 96.70%, correspondingly. In case early detection of CCs is performed, its treatment can be done with success and the survival rate can be improved. DLTs are popularly employed for automating detection of CCs, since they detect CCs with better accuracies. In this work, the Screening of CCs was addressed by the application of DLTs.

From the above, conclude that there are unsatisfactory results for the detection of abnormal subjects and the existing works focused on single MRI modality, which might have inherent limitations in prognostic prediction.

Radiomics is an evolving field that refers to a process that extracts high-throughput quantitative features from medical images, such as CT, ultrasound and MRI and builds discriminative models to assist in clinical decision-making. Recent reports have highlighted that convolutional neural network (CNN)-based radiomic methods are capable of learning the image features automatically, and the classification probabilities would be obtained via a softmax output layer, which is more convenient and more objective which is why this work uses CNN. Instead of the manual extraction of handcrafted features, the CNN with multiple layers allows the network to be fed with pixel values of an image, and learn the features with multiple levels of abstraction.

3. Proposed methodology

With the use of multi-parametric MRI data, this survey aims to propose models based on radiomic methods using DLTs for differentiating vessel invasions in CCs. DSSAMCNN develops the first two convolution layers using deep separable convolutions for reducing parameter counts and preventing model's data over fits. The library of MRI pictures includes both regular and irregular cervical images. These photos were taken from TCIA (The Cancer Imaging Archive). On the basis of 180 cervical MRI images, the dataset used in this proposed study is divided into three categories: normal, benign and malignant. About 90 of these were used for the training stage, and the remaining 90 were used for the testing phase during the training and evaluation of DLTs. Utilizing the ROCs (receiver operating characteristics) curve and confusion matrix analysis, predictive performances were evaluated. Maximum average areas were obtained using attention ensembles in learning where both MRI sequences were joined. The resultant outcomes were contrasted with other methods, demonstrating that radiomic methods using DLTscan be of

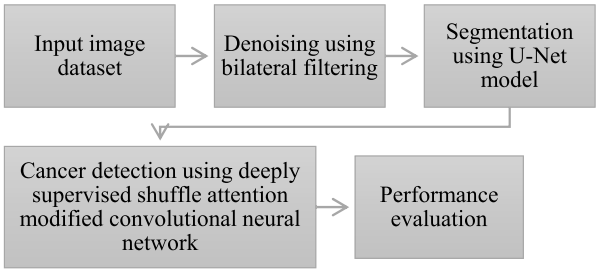

Figure 1. General framework diagram for detections of CCs.

great assistance to radiologists in preoperative stages for detections of vessel invasions in CCs and depicted as Figure 1.

3.1. Dataset details

U.S. National Cancer Institute's TCIA archive for pertinent data and cancer imaging. The data collected from about 37,568 people is represented by the 30.9 million radiological pictures held by TCIA. This information is organized and categorized according to the types of tumours, and certain collections additionally include analytical or clinical data. Before information is made available via a web browser or programmatic interfaces, TCIA personnel is careful in de-identifying and choosing all submitted collections. Digital Object Identifiers, which serve as a point of reference for each released collection inside TCIA, are assigned to each collection. Additionally, by requesting a TCIA-generated Digital Object Identifier, researchers using the TCIA data may make the portion of the data that they utilize in their study publicly available. This data description examines a selected portion of the TCIA datasets that are already available to the general public. It explains the curation and publishing methods employed by TCIA and makes 15 collections of cancer imaging data available.

3.2. Bilateral filter-based noise removals

Image denoizing with edges are maintained for efficiency. This work proposes iterative bilateral filters to minimize Rician noises in MRI images where these filters increase denoizing efficiencies while preserving minute structural details and reducing biases caused by Rician noises. Bilateral filters are commonly defined by the variance values of the two Gaussians that determine the weights. This results in maintaining diagnostic qualities and visualizations of images. The idea of behavioural filtering is to operate within an image's dynamic ranges just like traditional filter operation [15]. Two pixels can be placed in close proximity to one another (either by positioning them in nearby spatial positions or assigning them close values), most likely in a perceivably effective way. The encompassing region in the domain is indicated by closeness, while the enclosing region in the limits is specified by similarity. By balancing the pixel values with coefficients that decrease in value as distance increases, conventional filtering, also known as domain filtering, places an emphasis on proximity. Range filtering also offers average picture values with weights decomposing with differences. The domain filter and the range filter are the two filter kernels that make up the bilateral filter kernel found in every neighbourhood. The domain filter's weights are proportional to how far a pixel is from its immediate surroundings in space. The radiometric distance surrounding a pixel is proportional to the range filter coefficients. Using, the bilateral filter's response at pixel point x is expressed as

$$\hat{I}(x,y) = \frac{1}{NR} \sum_{y \in NR(x)} d_s(x,y) * r_r(x,y) * I(y), \quad (1)$$

$$NR = \sum_{y \in NR(x)} d_s(x,y) * r_r(x,y), \quad (2)$$

where $NR(x)$ indicates the neighbour hood region surrounding x, y refers to the position in the neighbourhood, d_s and r_r signifies the domain and radiometric components of the bilateral filter, correspondingly, and are expressed as below:

$$d_s(x,y) = exp\left(\frac{-|x-y|^2}{2\Sigma_s^2}\right), \quad (3)$$

$$r_r(x,y) = exp\left(\frac{-|\text{int}_x - \text{int}_y|^2}{2\Sigma_r^2}\right). \quad (4)$$

The weight function r_s reduces with increases in spatial distances between x and y, and function r_r reduces with increases in radiometric distances between intensities int_x and int_y. The bilateral filter is influenced by three parameters, N – the support of the filter where the bigger value of N generates increased smoothing, Σ_s and Σ_r regulates decays of two weight factors. But, there is no theoretical evidence on the choice of the optimal values for Σ_s and Σ_r. However in [16], it is found that the optimal Σ_s has comparatively no sensitivity to noise; But, Σ_r has a nearly linear relation with the actual noise standard deviation Σ, it is also recommended that a reasonable range for σ_s must be in the limit. Let $O_1, O_2, \ldots, O_n$ refer to n statistically independent observations O from an area of fixed MRI image signal intensity SI [17]. So the joint PDFs (Probability Distribution Functions) of the observations can be expressed as:

$$PDF(\{O_i|SI\}) = \prod_{i=1}^{n} \frac{MI_i}{\Sigma^2} e^{\frac{MI_i^2+SI^2}{2\sigma^2}} I_0\left(\frac{SIM_i}{\Sigma^2}\right) \quad (5)$$

Provided the observed data and MI a model of interest, the unlabelled parameters in the PDF can be predicted through the maximization of the respective likelihood

functions LF(SI) or equivalently ln LF(SI):

$$\ln LF(SI) = \sum_{i=1}^{n} \ln\left(\frac{MI_i}{\Sigma^2}\right) - \sum_{i=1}^{n} \frac{MI_i^2 + SI^2}{2\sigma^2} + \sum_{i=1}^{n} \ln\left(\frac{SI * MI_i}{\sigma^2}\right) \quad (6)$$

The *ML* estimate is later got from the global maximum of ln LF w.r.t SI and Σ^2:

$$\{\widehat{SI}_{M*LF}, \hat{\Sigma}^2_{M*LF}\} = \arg\max_{SI,\Sigma^2}\{\ln LF\} \quad (7)$$

In order to estimate the signal and noise variance, this technique assumes that the underlying signal is fixed within the immediate vicinity. If the majority of the nearby neighbourhoods support this assumption, the mode of all MLF-estimated local noise levels may be used as a trustworthy estimator for the noise variances:

$$\hat{\Sigma}^2 = mode\{\hat{\Sigma}^2_{MLF(i)}\} \quad (8)$$

where $\hat{\Sigma}^2_{M*LF}$ refers to the local estimate of the noise variance (depends on the Rician noise features applying ML estimation) at every pixel i in the image utilizing a small local neighbourhood. The advantage of this method is that noise estimate is not dependent on the Rayleigh dispersed background region (the techniques, which depend on image background cannot succeed with no background area present). It should be noted that the MLE must be determined numerically using optimization techniques as there is no closed form solution to get the values in Equation (8). Prior to applying the bilateral filter iteratively in this approach, the bias is reduced. This is dependent on the Rice distribution's second moment, where bias can be reduced by removing $2\sigma^2$ from the squared image. Let $\hat{I}_i$ refer to the denoised image post i iteration, so the bias in the image gets updated as:

$$\hat{I}_i = \sqrt{\max(I_i^2 - 2\hat{\Sigma}^2, 0)} \quad (9)$$

So the bilateral filter is used in an iterative manner and the noise level σ will decrease post every iteration and needs re-estimation post every iteration. The technique provided in Equation (9) is utilized for estimating the noise post every iteration. Next, the denoizing of the segmentations that the recommended model predicted was performed.

3.3. Detections of CCs using DSSAMCNN

In this section, the segmentation of CCs region is considered to be a multipixel classification problem, and the multimodal information of CCs image is exploited. Four different modalities of brain tumour pictures are provided to the network. Finally, the network will create a picture with the same dimensions as the input image, each pixel point having a label for a prediction category. Figure 2 provides an illustration of the DSSAMCNN's architecture. The network uses auxiliary outputs that are intensely supervised, densely linked convolutions and residual blocks. The subsequent sub-parts of this research provide a detailed explanation of the major network segments. A modified U-Net architecture is suggested, initially put out by Ronnebergeret al. [18], in the form of a backbone model. A contraction (encoder) segment and a localization (decoder) segment are both parts of the modified U-Net model. patches of input data with a size of $1 \times 160 \times 160$ ($\mathbb{C} \times \mathfrak{p}\mathrm{h} \times \mathrm{w}$) pixels with G groups, where $\mathbb{C}$ refers to the number of image channels, w the width and $\mathfrak{p}\mathrm{h}$ the patch height, were utilized as the input to the modified CNNs (Figure 2).

SA module's structure which splits input feature maps into several sets before combining channel and spatial attentions into single blocks for groups using Shuffle Units. Subsequently, sub-features are compiled and operators called "channel shuffles" are employed to make it easier for different sub-features to communicate information between them.

Channel attention: A choice for completing acquiring the channel-wise dependencies uses the densely connected convolutions block introduced in [19]. But, there will be excessive number parameters, which is undesirable during the design of a more flexible attention module in terms of a balance between speed and accuracy. In addition, it is not apt for generating channel weights by carrying out a rapid 1-D convolution of size k like Efficient Channel Attention Model [20] since$k$ intends to be bigger. For the improvement, a choice is provided, which initially incorporates the global information by just making use of GAP (global averaging pooling) for establishing channel-wise statistics as $s \in R^{\mathbb{C}/2G\times 1\times 1}$, computed by reducing X_{k1} through spatial dimension $\mathfrak{p}\mathrm{h} \times \mathrm{w}$:

$$s = \frac{1}{\mathfrak{p}\mathrm{h} \times \mathrm{w}} \sum_{i=1}^{\mathfrak{p}\mathrm{h}} \sum_{j=1}^{\mathrm{w}} X_{k1}(i.j) \quad (10)$$

The contraction unit of the modified U-Net is exploited for extracting there are many layers of picture feature delineations, and there are two sets of convolution blocks. With a 3×3 convolution kernel and deep separable convolution layers, each convolution block also has leaky rectified linear unit (Leaky ReLU) activations with negative slopes of 0.01, followed by batch normalizations [21] for numerical stability. Max Pooling layers with 22 kernels and stride of 2, to down samples, come before convolution blocks. Localization approach uses trilinear up sampling functions, in order to move the lower resolution, differentiative features into a higher resolution pixel space, each up sampling is performed before to a convolution block. The contracting path's characteristics are duplicated in the U-Net architecture

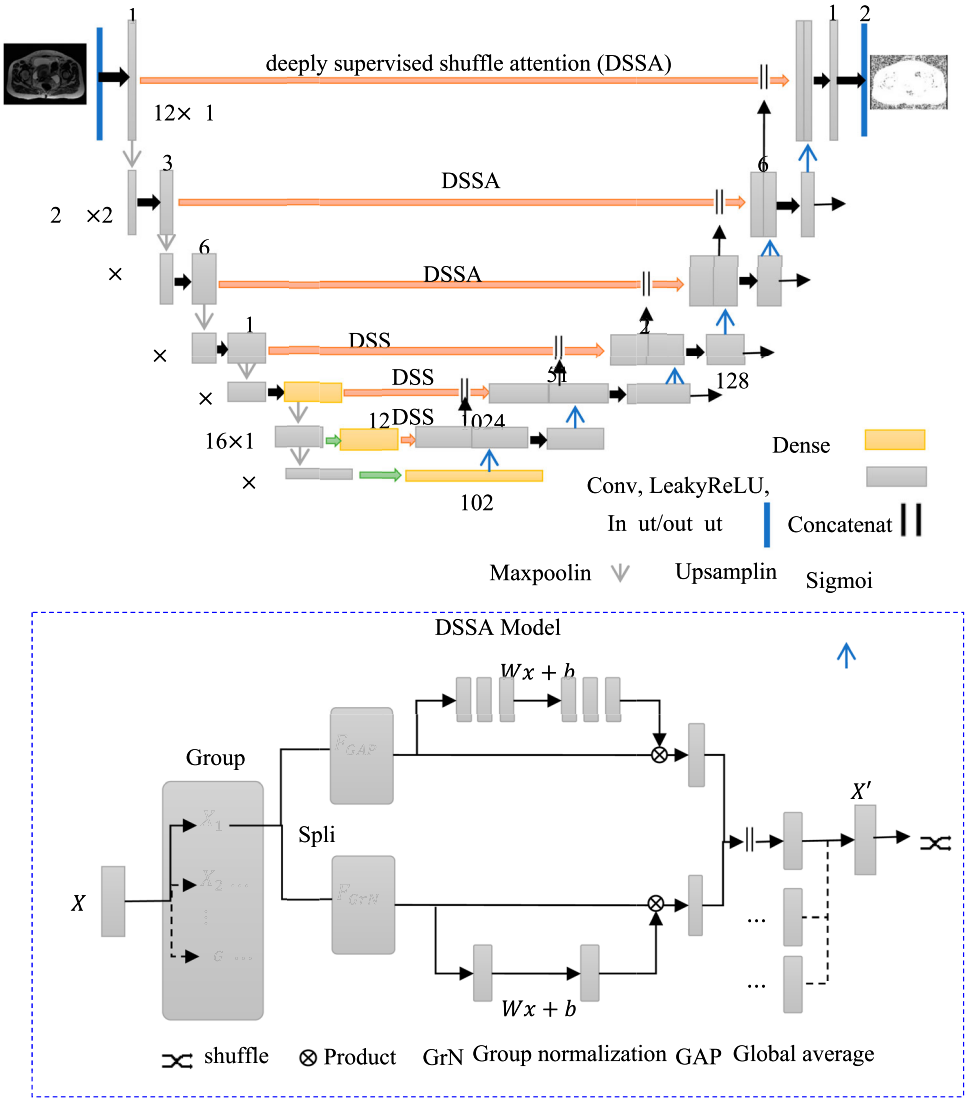

Figure 2. Proposed DSSAMCNN model for medical image segmentation.

and combined with the relevant localization path features via skip connections at various levels. The information that was lost as a result of the down sampling procedures carried out in the model's encoder section can be recovered using this. Deep supervision uses secondary segmentation maps with different resolutions that were created using the Pelvic U-Net architecture. These may be used to compute the weighted auxiliary losses, which are added to the principal loss function.

DSSAMCNN has SA modules that are concatenated as inspired by survey of Zhang and Yang [22] (Figure 2). The SA module provides the extra benefits of both spatial attention and channel-wise attention, emphasizing the key features and the potential places for these features. In the beginning, a SA module separates a feature map X into Gr groups, each of which represents a subset of X's features. The number of groups was determined to be fixed as $G = 16$ for this study as a hyperparameter. A branch for channel attention and a branch for spatial attention are once more created for each sub-feature. The channel attention branch uses the GAP, which comes before a sigmoid-activated attention mechanism. The spatial attention block employs group normalizations that precede the sigmoid activation. Concatenation is used to later combine the results of the two attention branches. In order to enable cross-information inflow, a channel shuffle approach described in ShuffleNet V2 [23] is finally used in channel dimensions.

The last layer of the contracting route produces a high-dimensional image feature with strong semantic information, and convolutions with tight connections are employed to remove the recurrent features that the network learns, as shown in Figure 2. The dense block has three convolution blocks, each of which has two convolutions of size 33. Utilized are batch normalization and the leaky ReLu activation function. Convolutions with tight connections imply that learning of the feature map from all the earlier connected feature maps is feasible, facilitating feature reuse and enhancing network segmentation performance. Through 5-fold cross-validation using the adaptive moment estimation

(Adam) optimizer, the model was trained [24], a learning rate of $lr = 0.01$ and a batch size of 2. The learning rate was reduced by applying a polynomial learning rate decay schedule having a power of 0.9. Every cross-validation model was trained for an overall of 1000 epochs, with 250 iterations for each epoch. 2D patches sized $1 \times 160 \times 160$ pixels ($\mathbb{C} \times \mathfrak{p}\mathrm{h} \times \mathrm{w}$) were randomly cut off from MRI image regions and sent to the input layer of the CNNs. To prevent overfitting during training, data augmentation approaches such as arbitrary rotations (±15°, nearly inferior-superior axis), mirroring (inferior–superior, left–right axis), scaling (0.85–1.25) and gamma augmentations (0.7–1.5) were utilized. The model was optimized by merging the soft Dice D and loss in BCEs (binary Cross-entropies) [25]:

$$Loss = Loss_D + Loss_{BCE} \tag{11}$$

$$Loss_D = -\frac{2\sum_{i=1}^{N} \hat{y}_i \cdot y_i + \omega}{\sum_{i=1}^{N} \hat{y}_i \sum_{i=1}^{N} y_i + \omega} \tag{12}$$

$$Loss_{BCE} = -\sum_{i=1}^{N} y_i \log(1 + \hat{y}_i + \omega) + (1 - y_i) \log(1 - \hat{y}_i + \omega) \tag{13}$$

where y_i indicates an annotated ground truth region, $\hat{y}_i$ the respective prediction and N stands for total images counts. Minimal constants $\omega = 1 \times 10^{-8}$ are included in both soft Dice losses and the BCE losses so that numerical stabilities are ensured. Segmentations map successive deep supervision phases that were exponentially weighed in loss functions of model's decoder paths, with lowest weighed decoder outputs having the lowest resolutions and highest weighed decoder outputs having the greatest resolutions:

$$\mathcal{W} = \sum_{i=1}^{\mathbb{S}-1} \frac{1}{2^i} \tag{14}$$

where $\mathbb{S}$ refers to the number of modified U-Net stages. All the weights were regularized to add to one. The final loss *FinalLoss*, with the exception of the lowest phase of the modified U-Net, is expressed by:

$$Final\ Loss = \sum_{i=1}^{N} \mathcal{W}_i [Loss]_i \tag{15}$$

4. Experimental results and discussion

This section examines the outcomes and comparisons of the experiments. All of the tests for this survey were run on NVidia GeForce GTX 1080 Ti graphics cards with 12 GB of memory and DLTs and Windows 10 computers with an Intel Core i7 CPU and 16 GB of RAM. Ten models were trained for 10 cross-validation sets in order to illustrate the approach in an objective and reliable manner. The DSSAMCNN in this study was trained over 400 iterations using stochastic gradient descent with a momentum of 0.9, learning rate of 0.001, batch size of 4, decaying weights of 0.0001 and gradients cropped to 5.0 in iterations. This work tested with lengthy training durations and several training hyper-parameter combinations, however, very limited improvements and curve losses post 300 training epochs were smoothened. Figure 3 shows the results of proposed detections of CCs using DSSAMCNN-based segmentation.

In this method, models trained with 400 epochs for coarse segmentation are employed in this survey. All of the cell pictures' shorter sides were resized to 512 pixels in length in order to reduce image compression and adjust the feature extraction network. The segmented image will then be given as input to the proposed DSSAMCNN to detect the cervical contour, and various parameters will be used to validate the proposal's performance in terms of DSCs, JSIs (Jaccard Indices), accuracy, execution time, specificity and sensitivity. Finally, comparisons with other methods such as K-means MVFA-Net and U-Net and accuracy are done in this work. When measuring DSCs, the following equation is used to calculate the actual percentage of the actual tumour or lesion and accessible nontumour or non-lesion pixels to the predicted tumour or lesion and non-tumour pixels:

$$\mathrm{DSC} = \frac{(2\mathrm{TP})}{(\mathrm{FP} + 2\mathrm{TP} + \mathrm{FN})} \times 100 \tag{16}$$

JSIs were taken into account when computing the degree of similarity between the number of actual tumour or lesion pixels taken into account and the

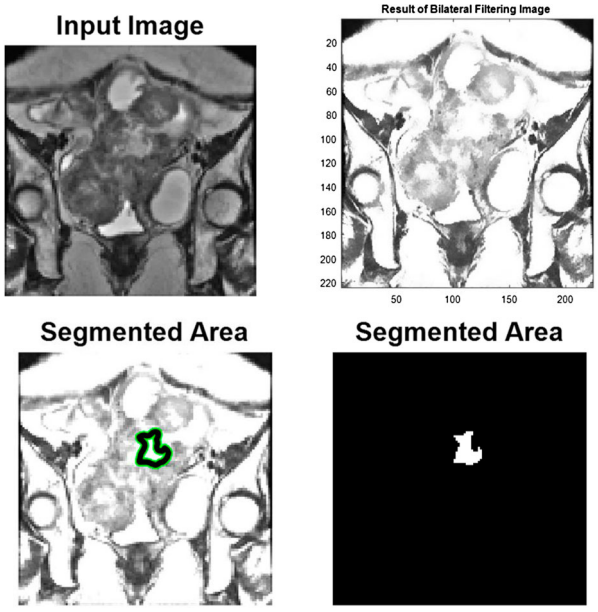

Figure 3. The results of proposed detections of CCs using DSSAMCNN-based segmentation.

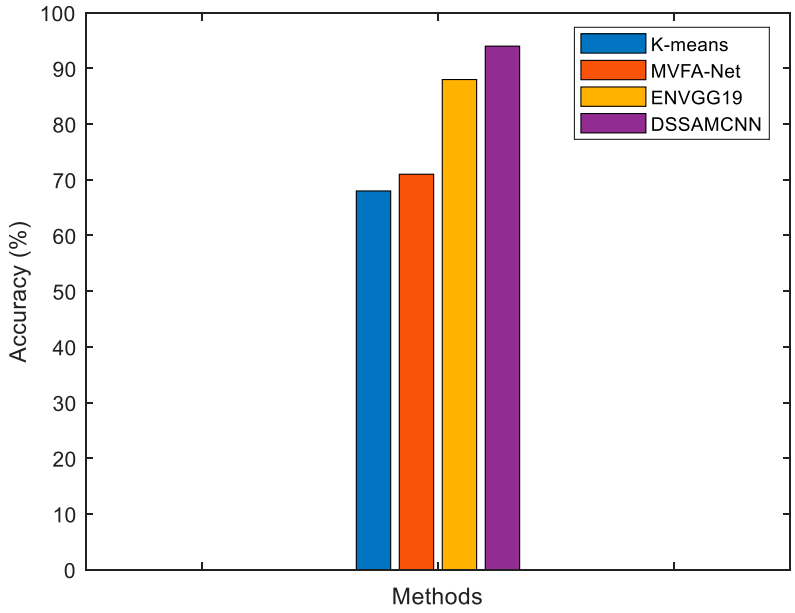

Figure 4. Accuracy performance comparison.

number of predicted tumour pixels.

$$\text{JSI} = \frac{(\text{TP})}{(\text{TP} + \text{FN} + \text{FP})} \times 100 \tag{17}$$

Accuracy is a crucial factor for determining the proper tumour or lesion region of interest classification rate, which is calculated using the equation below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN}) + (\text{FP} + \text{FN})} \times 100 \tag{18}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} * 100 \tag{19}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100 \tag{20}$$

In these phrases, the result is referred to as a TP (true positive) if the model accurately forecasts the positive class. The abbreviation FP stands for false positive and denotes a result that the model has estimated to be positive, whereas TN stands for true negative and denotes a result that the model had predicted to be negative. The model is referred to as FN if it incorrectly guesses the negative class (false negative). The numerical outcomes of the suggested and accessible procedures are listed in Table 1.

Figure 4 illustrates the accuracy achieved with the proposed and available models in terms of the number of features in a particular database. The proposed DSSAMCNN improves the accuracy yielding 94% and the existing methods such as K-means, MVFA-Net, U-Net produces only 68%, 71% and 88%. In comparison with the standard K-means, MVFA-Net, U-Net, the efficiency of the proposed technique DSSAMCNN is due to two important aspects. One is that the proposed deep model makes the best use of the benefits of the standard K-means, MVFA-Net and U-Net. But, in spite of the enhanced image contrast, differentiating between big and small bowel can be hard even for a skilled radiation oncologist, and it might also be the reason behind the slight inter-observer differences in the observational ratings.

Table 1. The numerical results of proposed and available techniques.

Metrics	K-means	MVFA-Net	ENVGG19	DSSAMCNN
Accuracy	68	71	88	94
DSC	73	76	77	91
Sensitivity	75	77	80	95
Specificity	78	80	85	97
JSI	68	73	78	88

Figure 5 shows that the DSSAMCNN model performs better than the common K-means, MVFA-Net and U-Net design in terms of DSC. This DSSAMCNNmodel yields a DSC of 73% on the test set model, which is higher than the available deep models by close to 1–2% while the other models such as K-means, MVFA-Net and U-Net produce 73%,76% and 77% accordingly. The inclusion of permissible noise during the learning of a task helps in improving the models, which is similar to eliminating the unwanted steps from the procedure. Contrary to traditional learning, interaction can help in avoiding the overfitting of various image scenarios from degradation. Consequently, the proposed DSSAMCNN learning approach can help in achieving several, comparatively relevant goals. In the case of all disease-infected region segmentation, the right epochs were utilized for optimizing the training time in the processapplyingDSSA. This has led to a quick and consistent fall in the DSC of the recognition model.

Figure 6 shows the JSI comparsion results. The proposed DSSAMCNN model, learned to identify the infected area in both skin images and fetal images,

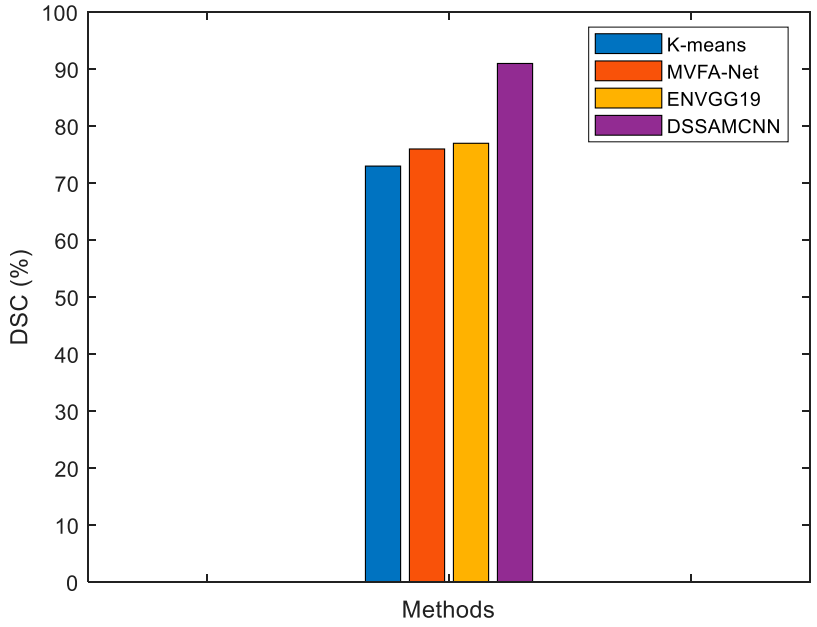

Figure 5. DSC comparison results.

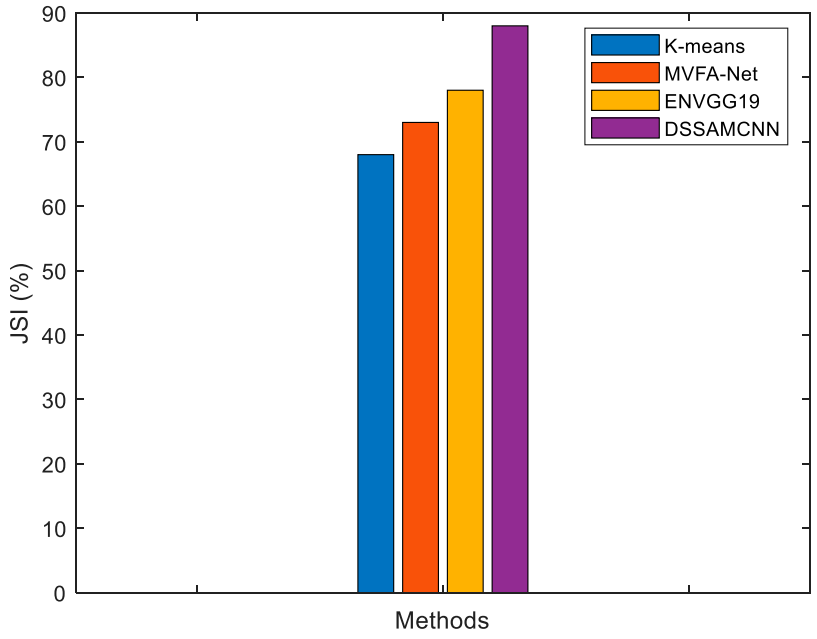

Figure 6. JSI comparison results.

depicted a dependable effect, which might be utilized in the form of a real smart technology in medical imaging. Other research explored into the uniqueness of the DL-based models of being capable of differentiating between diseases from various time slots. Figure 5 provides the comparison between the K-means, MVFA-Net, U-Net and DSSAMCNN. For the detection, recognition and description ofCCs, severalDLTs are utilized. DSSAMCNNachieves88% while the other existing methods such as K-means, MVFA-Net and U-Net methods give 68%,73% and 78% only. Moreover, DSSAMCNNtechnique is a rapid technique, whose computational complexity is less compared to other identical techniques in the segmentation process.

Figure 7 shows the specificity of proposed and available models in terms of the number of features in a particular database. With the rise in the number of features, there is also an increase in recall. The DSSAM-CNN yields a recall of 97% while the other models such as K-means, MVFA-Net and U-Net give 78%, 80% and 85%. Since the available techniques are plain models not sufficiently effective for high-dimensional datasets, they do underfit. It can be observed that not just the segmentation accuracy is improved; the stability and generality of the proposed modified U-Net algorithm are also reliable.

Figure 8 depicts the sensitivity of proposed and available models in terms of the number of features in a

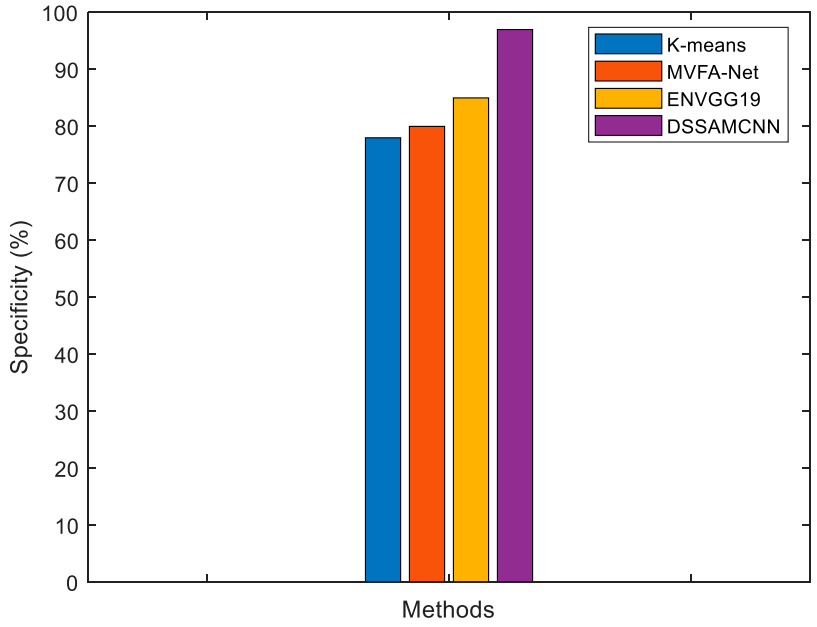

Figure 7. Specificity performance comparison.

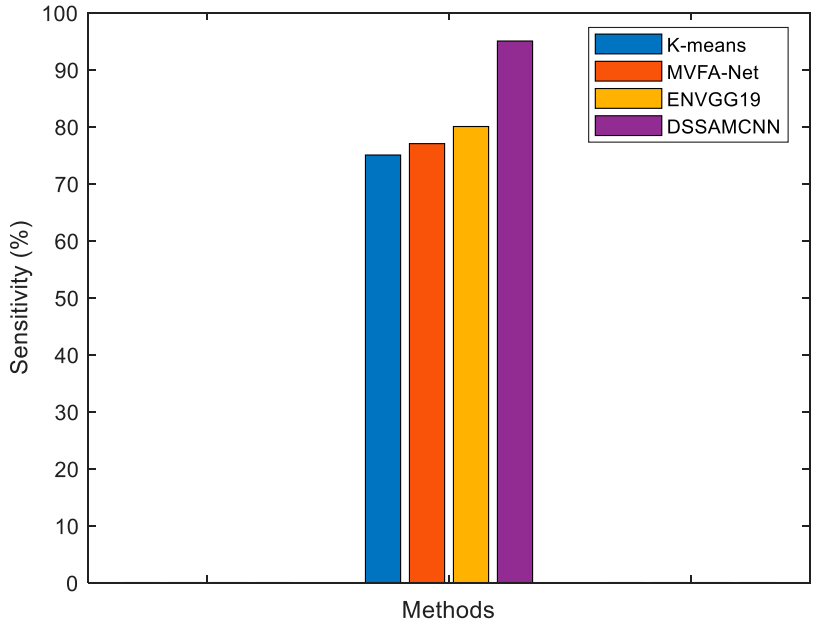

Figure 8. Sensitivity performance comparison.

particular database. The DSSAMCNN improves the accuracy and yields the sensitivity of 95% and the other existing methods k-means, MVFA-Net and U-Net produce 75%, 77%, and 80%. Therefore, the proposed algorithm is much better than the available algorithms in terms of improved validation outcomes for cancer prediction. Also, it is worth mentioning that the time required for the CNNs to have convergence is quite moreover other techniques. As a result, this is because of the structural depth of the DSSAMCNN, which generally needs a longer time, especially, if the number of inputs is huge. However, this deep structure constitutes the primary aspect in yielding a better identification rate in comparison with other networks.

5. Conclusion and future work

Cervical nuclear segmentation has a prominent part to play in automated cytology screening and diagnosis for CCs. In this research, an innovative CCs segmentation technique is introduced in which DSSAMCNNis merged with a DSSA. This approach is about sufficient conjunction and usage of historical and spatial information. The DSSAMCNN typically renders accuracy in boundary localization, primarily due to the reliable semantic information acquired under the presence of the pixel-level existing information, and the MCNN examines the spatial information, which includes the position, intensity and coarse segmentation results

attained with the help of DSSAMCNN in the nuclear cervical MRI images, for clarifying the nuclear boundary. It is confirmed from the results of the experiments carried out on the TCIA dataset that the proposed technique is quite efficient in providing nuclear segmentation much better than that can be achieved with the other approaches. This work can be trusted for the results being trustworthy for the next-level analyses of automated MRI screening and diagnosis for CCs. Results show that the proposed model achieves 94% accuracy, 91% DSC, 95% Sensitivity, 97% Specificity and 88% JSI. Even though this work has made considerable improvement in the cervical nucleus segmentation, there is still room for improving the accuracy achieved with the uneven nuclear segmentation due to its clinical importance. Also, the present work can have an easy extension to MRI screening of other type of cancers. As a futuristic approach, the spatial and channel attention modules of SA need to be explored and then use them in many upcoming CNNs architectures, such as the ShuffleNet family, the SKNet and MobileNetV3.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Cohen PA, Jhingran A, Oaknin A, et al. Cervical cancer. Lancet. 2019;393(10167):169–182.

[2] Bobdey S, Sathwara J, Jain A, et al. Burden of cervical cancer and role of screening in India. Indian J Med Paediatr Oncol. 2016;37(04):278–285.

[3] Dappa E, Elger T, Hasenburg A, et al. The value of advanced MRI techniques in the assessment of cervical cancer: a review. Insights Imaging. 2017;8(5):471–481.

[4] Adweb KMA, Cavus N, Sekeroglu B. Cervical cancer diagnosis using very deep networks over different activation functions. IEEE Access. 2021;9:46612–46625.

[5] Sartor H, Minarik D, Enqvist O, et al. Auto-segmenta tions by convolution neural network in cervical and anorectal cancer with clinical structure sets as the ground truth. Clin Translational Radiat Oncol. 2020; 25:37–45.

[6] Devi MA, Sheeba JI, Joseph KS. Neutrosophic graph cut-based segmentation scheme for efficient cervical cancer detection. J King Saud University – Comput Inf Sci. 2022;34(1):1352–1360.

[7] Rigaud B, Anderson BM, Zhiqian HY, et al. Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer. Int J Radiat Oncol Biol Phys. 2021;109(4):1096–1110.

[8] Gou S, Xu Y, Yang H, et al. Automated cervical tumor segmentation on MR images using multi-view feature attention network. Biomed Signal Process Control. 2022;77:103832.

[9] Lu P, Fang F, Zhang H, et al. AugMS-Net: augmented multiscale network for small cervical tumor segmentation from MRI volumes. Comput Biol Med. 2022;141:104774.

[10] Rajarao C, Singh RP. Improved normalized graph cut with generalized data for enhanced segmentation in cervical cancer detection. Evol Intell. 2020;13(1): 3–8.

[11] Liang P, Sun G, Wei S. Application of deep learning algorithm in cervical cancer MRI image segmentation based on wireless sensor. J Med Syst. 2019;43(6): 1–7.

[12] Lin YC, Lin CH, Lu HY, et al. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. Eur Radiol. 2020;30(3):1297–1305.

[13] Lakshmi GA, Ravi S. Automated segmentation algori thm for cervical cell images by employing cuckoo search based ICM. J Ambient Intell Humaniz Comput. 2017: 1–11.

[14] Bai B, Liu PZ, Du YZ, et al. Automatic segmentation of cervical region in colposcopic images using k-means. Australas Phys Eng Sci Med. 2018;41(4):1077–1085.

[15] Bhonsle D, Chandra V, Sinha GR. Medical image denoising using bilateral filter. Int J Image Graph Signal Process. 2012;4(6):36.

[16] Park J, Han JH, Lee BU. Performance of bilateral filtering on Gaussian noise. J Electron Imaging. 2014;23(4):043024.

[17] Riji R, Rajan J, Sijbers J, et al. Iterative bilateral filter for Rician noise reduction in MR images. Signal Image Video Process. 2015;9(7):1543–1548.

[18] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention (pp. 234–241). Springer, Cham. 2015, October

[19] Ali Z, Irtaza A, Maqsood M. An efficient U-Net framework for lung nodule detection using densely connected dilated convolutions. J Supercomput. 2022;78(2):1602–1623.

[20] Bian J, Liu Y. Dual channel attention networks. J Phys: Conf Ser. 2020, September;1642(1):012004, IOP Publishing.

[21] Gitman I, Ginsburg B. Comparison of batch normalization and weight normalization algorithms for the large-scale image classification. arXiv preprint arXiv. 2017;1709:08145.

[22] Zhang QL, Yang YB. Sa-net: shuffle attention for deep convolution neural networks. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2235–2239). IEEE. 2021, June

[23] Ma N, Zhang X, Zheng HT, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design. Proceedings of the European conference on computer vision (ECCV) (pp. 116–131). 2018

[24] Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv. 2014;1412:6980.

[25] Janthakal S, Hosalli G. A binary cross entropy U-net based lesion segmentation of granular parakeratosis. In 2021 international Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1–7). IEEE. 2021, October