



OPEN

# Classification of cervical cancer using Dense CapsNet with Seg-UNet and denoising autoencoders

Hui Yang<sup>1</sup>✉, Walid Aydi<sup>2,3</sup>, Nisreen Innab<sup>4</sup>, Mohamed E. Ghoneim<sup>5,6</sup> & Massimiliano Ferrara<sup>7</sup>✉

Cervical cancer is one of the deadly diseases that affects women, which requires periodic examinations to identify and treat any cancerous tumors at a preliminary stage. The most prevalent examination tool for cervical cancer prompt identification is the cervical smear (Pap smear) testing; however, due to human negligence, this examination method has an elevated probability of negative findings. Cervical cancer classification using machine learning (ML) and deep learning (DL) has been extensively studied to enhance the conventional diagnostic process. Robust classification results were achieved through the pre-segmented imagery in most current investigations. Conversely, cellular grouping makes reliable cervical cellular segmentation difficult. Additionally, the deep learning methods used in the existing works perform poorly on a multiclass classification when the data distribution is skewed, which is common in the cervical cancer dataset. To mitigate these restrictions in cervical cancer research, this proposed work uses a combination of four different deep-learning methods in various phases of this research. The proposed work is segregated into five phases: pre-processing, data augmentation, segmentation, feature extraction, and classification. Contrast maximization is performed in the pre-processing phase, and the images are augmented using Multi-modal Generative Adversarial Networks (m-GAN) in the second phase. In the third phase, cervical cancer images are segmented using the Seg-UNet model, which is forwarded to the feature extraction phase that employs denoising autoencoders. Finally, the classification is implemented using the Dense CapsNet model and applied to the SIPaKMeD dataset to categorize between normal, abnormal, and benign classes. The proposed system achieves an accuracy of 99.65%, which is higher than the other works in the literature.

**Keywords** Cervical cancer, Deep learning, Segmentation, Autoencoders, Dense CapsNet

The fourth-leading cause of cancer among women is cancer of the cervical cavity, which appears in the female reproductive organ of the woman. The World Health Assembly of the United Nations estimates that 0.36 million women died from cervical cancer in 2021 solely, out of an estimated 0.59 million women who received an identification of the deadly condition worldwide<sup>1</sup>. Due to a lack of examination and therapy amenities, more than seventy percent of cervical carcinoma incidences and seventy-five percent of casualties occur in insufficiently developed and economically deprived nations<sup>2</sup>. Premature gestation, tobacco use, inadequate sanitation during menstruation, and the use of dental appliances are among the primary risk factors for acquiring the virus that causes human papilloma. Cancer of the cervical region is mainly brought on by this persistent infection, according to research. However, if caught early and given the proper care, cervical cancer is a relatively curable type of cancer<sup>3</sup>.

Female cervixes are covered in a fragile film of cell-based tissues. Cervical cancer is the term used to describe a condition in which a cell transforms into a carcinogenic tumor with the ability to multiply and spread quickly<sup>4</sup>.

<sup>1</sup>Department of Critical Medicine, Baoshan People's Hospital, Baoshan 678000, Yunnan Province, China.

<sup>2</sup>Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, 11942 Al-Kharj, Saudi Arabia. <sup>3</sup>Laboratory of Electronics & Information Technologies, Sfax University, Sfax, Tunisia. <sup>4</sup>Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Diriyah, 13713 Riyadh, Saudi Arabia. <sup>5</sup>Faculty of Computers and Artificial Intelligence, Damietta University, Damietta, Egypt. <sup>6</sup>Mathematics Department, Faculty of Sciences, Umm Al-Qura University, Mecca, Kingdom of Saudi Arabia. <sup>7</sup>Decisions LAB, Department of Law, Economics and Human Sciences, University Mediterranea of Reggio Calabria, Via dei Bianchi, 2, 89131 Reggio Calabria, Italy. ✉email: huiyangscientif@outlook.com; massimiliano.ferrara@unirc.it

Usually, a sampling and an imaging test are used to determine the illness. Determining the cancer's spread is possible using computational image analysis tools. For enabling timely diagnosis and therapy, periodic examination of women who are around thirty years of age is essential for accurately minimizing cervical cancer<sup>5</sup>.

Cervical cellular pathology is the most frequently employed diagnostic method for cancer of the cervical area because of its reasonable price<sup>6</sup>. Using this method, specialized cellular biologists remove cells from the squamocolumnar end of the uterus and examine them under a fluorescent microscope to determine whether they are carcinogenic. The analysis of one sheet typically takes approximately five to ten minutes to evaluate due to the various cellular configurations and duplication<sup>7</sup>. Additionally, every image contains roughly 3,000,000 cells in order with varying positions and crosses over, making the conventional assessment method difficult, tedious, exorbitant, and highly susceptible to inaccuracies. This leads to developing a controlled electronic detection system that can reliably and successfully analyze the test for cervical cancer cells<sup>8</sup>.

A lot of research has been done to create an automated evaluation system that would assist physicians in diagnosing cervical tumors because information regarding training became available around the middle of the early nineties<sup>9</sup>. Three types of operations make up a typical automated system, including the extraction of features, categorization, and cellular division into cell membranes and mitochondria. Initially, preliminary processing based on filter techniques is carried out in this system to improve the quality of imagery. Then, using strategies for clustering, the mitochondria of cells are retrieved<sup>10</sup>. The segmented cytoplasm is subsequently rectified through additional processing. The separated nuclei are then used to extract customized elements, including structural, hue measurement, and texture-related characteristics. In the meantime, the process of choosing characteristics is employed to identify the most discerning traits, and ultimately, a system for classification is developed to categorize the cell components<sup>11</sup>.

The previously outlined approach demonstrates the inadequacy of computerized learning by requiring numerous steps to examine the data and extract customized features that cannot guarantee higher classification accuracy<sup>12</sup>. Deep learning (DL) based methods for extracting features offer a significant distinction from other machine learning (ML) algorithms in terms of producing a more effective automated system<sup>31</sup>. The biggest growing field of computational intelligence, deep learning<sup>32</sup>, has found profitable uses in a wide range of industries, including physiological visualization<sup>33</sup>, self-driving cars, military operations, the farming industry, and healthcare<sup>13</sup>.

Deep learning methods extract wider, preliminary, and exceptional characteristics from imagery using numerous layers of information, thereby learning structural aspects. Reliable visual identification, differentiation, and categorization are potential applications of deep learning. The analysis of natural languages has progressed substantially in the past few decades owing to sophisticated deep-learning methods<sup>14</sup>. These approaches have been incredibly beneficial in various healthcare industries, including tomography, cytology, and ultrasound, particularly in identifying and treating malignancies. Using imagery gathered from various healthcare visualization methodologies, Convolutional neural networks, and perception conversion models have demonstrated favorable tumor recognition and prognosis results.

A potential drawback of deep learning-based techniques is that, in contrast to machine learning methods, they require enormous information to produce favorable outcomes, which might be difficult to gather in healthcare. Furthermore, this is a common occurrence in the medical arena; DL performs poorly in multiclass analysis categorization issues where the sample data is distributed inconsistently<sup>15</sup>. Additionally, the absence of information with labels required to calibrate models is another issue with DL models. Another challenge is the high degree of fluctuation between the observations made by physicians, which arises from the possibility of differing specialist assessments of an identical histological image<sup>16</sup>. Additionally, deep learning models can overgeneralize the training set, leading to substandard adaptation on untrained data. Therefore, more investigation and development are needed for the automated technique based on deep learning adopted for evaluating cells that cause tumors in cervical regions.

### Motivations of current research

The current research is conducted to answer the following research questions related to cervical cancer classification.

RQ1: Which preprocessing methods are best for improving the coherence and standard of imagery related to cervical cancer?

RQ2: Which deep learning architectures are best suited for classifying cervical cancer cases, considering variables like computational performance, comprehensibility, and model intricacy?

RQ3: How can we create novel architectures specifically engineered to use the distinct features found in images of cervical cancer, like atypical cell shapes and different degrees of lesion severity?

RQ4: In what ways could the optimal design of data augmentation techniques produce a variety of synthetic samples that enhance the resilience and applicability of deep learning models for the classification of cervical cancer?

RQ5: In terms of cervical cancer categorization, how do deep learning models vary compared to current diagnostic techniques?

### Research contributions

The main contributions of this paper are,

1. To employ Multi-modal Generative Adversarial Networks (m-GAN) for data augmentation of images of the tumors in the cervical cavity.
2. The Seg-UNet model will be incorporated to segment the pre-processed images, and denoising autoencoders will be utilized to perform feature extraction.

3. To leverage the Dense CapsNet model for three class and five class classification of cervical cancer using the SIPaKMeD dataset.

## Paper organization

Section “[Related works](#)” presents the latest advancements in cervical cancer detection using machine learning and deep learning methods. Section “[Proposed methodology](#)” discusses the proposed methodology involving five phases for classifying cervical cancer. Section “[Results and discussion](#)” explains the dataset used in the experimentation and compares the outcomes obtained for 3 and 5 class classifications. Section “[Conclusion](#)” concludes the present research.

## Related works

This section presents the current developments and investigations conducted in cervical cancer detection using machine learning and deep learning approaches. Presents an approach that utilizes data from the imagery of the cervical regions<sup>17</sup>. The authors utilized an improved contour box technique to identify the region of interest in cervical pictures. Typically, the region is a contour box form with the cervical canal at its middle point. Using an analogous measurement, multiple contour areas with various orientations and dimensions can be identified from associated imagery. Next, a mixture of Manhattan distance and overlap over combination criteria is used to preserve the most desirable contour region. A binary classification model is constructed using the random forests model and the average polling method. Features that are developed or customized are used to train the models. Descriptions of cervical hues and surfaces are used to build these characteristics.

The authors conducted an evaluation and analysis of numerous Artificial Neural Network designs utilized in cervical carcinoma categorization<sup>18</sup>. The most effective topologies are proposed for the dual categorization of the prevalence of cervical imagery as either healthy or detrimental cervical tissues. The reliability of these systems is compared with conventional detection methods such as PAP examinations and solution-based pathology diagnostics.

The authors in<sup>19</sup> suggested a localization-free cervical carcinoma predictor to address the difficulties associated with segmenting imagery. Unlike earlier models, the authors trained an algorithm using deep learning with automatically generated features. They used a multiple-stage methodology for training their model to accommodate the restricted volume of data required for training. The model developed using deep learning is trained beforehand on a collection of actual imagery in the initial stage. Regions with nuclei in the central part taken from cervical tissue images that have been continually reconfigured are then used to fine-tune the learning model. This model requires precisely generated picture areas despite its outstanding reliability.

The image of cervical spots, in which the cell membrane region is divided into eight parts, is considered by the deep learning classifier proposed in<sup>20</sup>. After that, these portions are divided into four structures that are classified using a deep neural network model. The parts and the entirety of the endometrium are categorized afterward by integrating the assessments obtained. The researchers evaluated their model with a restricted set of 67 cervical images, reaching an accuracy of 85%.

The investigators in<sup>21</sup> use a combination of deep learning techniques to expand on the work first proposed in<sup>22</sup>. For research purposes, morphological characteristics are taken from twelve horizontal units in the area of the endometrium. Twenty-nine synthetic and manually created characteristics are used to extract characteristics. A multilayer perceptron model built on windows with sliding mechanisms connected to each upward block is used to get the subsequent characteristics. The combined model classified the malignancy grade of the entire cellular area with an accuracy of 86.89%.

The research scientists in<sup>23</sup> suggested a small, productive neural network model compatible with devices with embedded chips and requires a few computations and variables. The research investigates how information reduction can improve the capacity of flexible neural networks for visualization. The research also discusses how essential the simulation selection is for the recommended strategy. The authors introduced the deep learning-based cervical carcinoma detection system in<sup>24</sup> to dynamically contribute to cervical cell classification procedures. The system uses a feed-forward neural network module for regional extraction of characteristics, and for regional characteristic retrieval, it uses an optical inverter element. Additionally, incorporating regional and globalized features for ultimate classification is designed through a back propagation neural network module. The suggested methodology aims to evaluate cervical images consistently and efficiently.

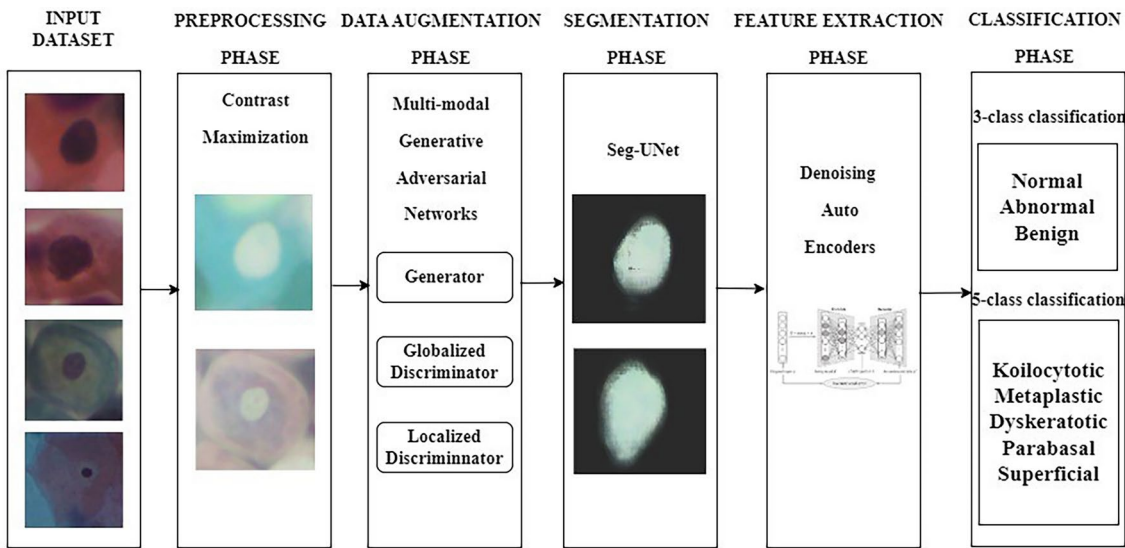
A technique for predicting cervical cancer based on the Naive Bayes classifier was presented in<sup>25</sup>. They used a portion of the database<sup>26</sup>, which included 156 cellular images, 104 of which were healthy and 52 of which were malignant. The accuracy of the system was 95%. Researchers cannot ascertain the reliability of the methodology in its entirety because the number of samples was limited. A technique for multilevel localization was used in<sup>27</sup> to segment images of cells in the cervical cavity. Although the degree of segmentation precision was 95.9%, some nuclear entities could not be identified owing to their chaotic appearance. In<sup>28</sup>, a stacked grouping was employed to carry out the categorization, along with another modular delineation technique. Due to multiple layers, the proposed method had issues with certain intracellular regions in the cervical cavity.

A cervical cell separation and categorization method was proposed in<sup>29</sup>. The authors combined Gaussian attributes, contour, and appearance characteristics for segmentation. For the distinction between healthy and tumor-like cells, the accuracy was higher than 92%. In<sup>30</sup>, a particle swarm optimization technique based on bio-inspired algorithms is employed for cervical cancer staging. One drawback of this technique is that the PSO endured irrational transformation, generating disproportionate solutions.

In this work, a new deep learning and explainable AI paradigm is proposed for the segmentation and classification of brain tumors, comprising the integration of DeepLabv3+ with a Bayesian Optimized for hyperparameter searching and two custom Inverted Residual Bottleneck architectures for classification. The research records distinct segmentation and classification accuracies at 92.68% and 95.42%, respectively, and

References	Database used	Segmentation	Classification	Inferences
Abdullah et al. <sup>8</sup>	Herlev dataset	Region-based segmentation	Multimodal deep network	The model is less complex
Li et al. <sup>9</sup>	UCI database	U-Net with Gaussian filtering	Ensemble model	Limited data
Miyagi et al. <sup>11</sup>	Microscopic images	Thresholding based segmentation	Stacked Autoencoder	Low accuracy and needs more investigation
Ples et al. <sup>12</sup>	Cervical smear images	U-Net	CNN-extreme learning machine	High Training time and model complexity
Hu et al. <sup>14</sup>	PAP smear images	Watershed transform	Random Forest and AdaBoost	Two-stage classification takes more execution time
Gaona et al. <sup>15</sup>	Colposcopy images	Fully automated segmentation method	DenseNet	Only suitable for colposcopy images
Bai et al. <sup>16</sup>	Microscopic images	3D U-Net	Inception V3 model	Suitable for binary classification problems only
Devi et al. <sup>17</sup>	Herlev dataset	Watershed transform	Fuzzy C-means algorithm	Poor classification results
Zhang et al. <sup>18</sup>	Cervical smear images	TransUNet	Fully convolutional neural network	Produces highly robust and accurate predictions of 96.5%
Skerrett et al. <sup>19</sup>	PAP smear images	V-Net	Pretrained ConvNet	Improper feature extraction affect the classification performance

**Table 1.** Comparison of existing research on cervical cancer classification.



**Fig. 1.** Proposed workflow.

combines LIME with the model internally for interpretability, thereby enabling transparency in the model. The outcomes point out the superiority of the described framework over conventional techniques and provide a leap in medical imaging by applying explainable AIs<sup>34,35</sup>. Table 1 provides a comparison of cervical cancer classifications.

**Proposed methodology**

The workflow of the proposed methodology involves five phases, as depicted in Fig. 1. The first phase includes preprocessing, which performs contrast maximization on the cervical cancer images<sup>2</sup>. In the second phase, data augmentation uses a Multi-modal GAN model, followed by segmentation in the third phase. The cervical tumor images are segmented by employing the Seg-UNet model, and these segmented images are passed on to the feature extraction phase. Denoising autoencoders are utilized in the fourth phase to extract the inherent characteristics in the images to facilitate efficient classification in the last phase<sup>5</sup>. The fifth phase includes the Dense CapsNet model, which performs multi-class classification of cervical cancer.

**Pre-processing phase**

Contrast Maximization is the first and most important stage in every computational imaging technique. The differences in the magnitude of the pixels between the diseased and benign zones are captured in this step. The region's scope must eventually be reduced for effective characteristics extraction and precise malignancy separation. As a result, contrast maximization is employed to modify the pixel strength of the cancerous regions within the appropriate intensity range. The luminosity of the affected areas is boosted using a reliable contrast-broadening method known as Pixel Expansion across Threshold, which is suggested for this purpose. The recommended approach increases the pixel's hue value by stepping up the specified parameters. This approach

comprises three steps: region selection, aggregating pixel values in selected regions, and determining the output through the Change Point Detection method.

To group the pixels that are identical to each other, three groups are created using the representations given in Eq. (1),

$$\begin{cases} G_1 : 0 \leq \theta(x, y) \leq 99 \\ G_2 : 100 \leq \theta(x, y) \leq 169 \\ G_3 : 170 \leq \theta(x, y) \leq 255 \end{cases} \quad (1)$$

Among these three groups,  $G_3$  is chosen to enhance the pixel strength by including an additional factor  $\delta$  with an arbitrary value of 20. If suppose the inclusion of this additional factor increases the strength of the pixel to be greater than 255, then the succeeding consecutive pixel is selected. This procedure is continually executed until all the pixels are computed as per Eq. (2),

$$P_s(x) = \begin{cases} \sum_0(\delta) & \text{if } G_3(x) \leq \gamma \\ \text{Otherwise} & \end{cases} \quad (2)$$

In order to improve the contrast of the malignant regions, limit functions are included as given in Eqs. (3) and (4).

$$\lim_{x \rightarrow 0} P_s(x) = \frac{1}{Z} \sum_{z=0}^{Z-1} H_z \quad (3)$$

$$\lim_{x \rightarrow 0} P_s(x) = H_{max} \quad (4)$$

$H_z$  in Eq. (3) denotes the standardized form of histogram that belongs to third group. The contrast maximized images  $\theta_z(max)$  computed using Eqs. (5) and (6) are saved into a separate collection and passed on to the next phase.

$$D_g = \sum_{g=0}^Z P_s \quad (5)$$

$$\theta_z(max) = (Z - 1) D_g \quad (6)$$

The results achieved in this pre-processing phase are presented in Fig. 2.

### Data augmentation phase

The structure and texture of the region in the cervical cavity are complex due to its cytological nature. Hence, it is a tedious process to generate the cervix images directly. In this work, m-GAN model is employed to implement the data augmentation of the cervical cancer imagery. This method creates new images by fusing the previously available images. For any image  $m$ ,  $m_i$  denotes the particular image, and  $l_i$  denotes the corresponding labeling. The generated image,  $l_{ij}$  is a concatenation of  $l_i$  and  $l_j$  of two different images  $i$  and  $j$ . The generator module of the m-GAN model is responsible for producing both the cervical cavity area and the tumorous part on it. Area-based loss function, as shown in Eqs. (7)–(9), is formulated to achieve this objective.

$$\begin{aligned} L_{da} = & \mu_1 H [\sigma_{1,2}(n^{heal}) - \sigma_{1,2}(m_j^{heal})]^2 \\ & + \mu_2 H [\sigma_{2,2}(n^{heal}) - \sigma_{2,2}(m_j^{heal})]^2 \\ & + \mu_3 H [\sigma_{2,2}(n^{can}) - \sigma_{2,2}(m^{can})]^2 \end{aligned} \quad (7)$$

$$n^{heal} = n \cdot A_j^{heal} \quad (8)$$

$$n^{can} = n \cdot A_i^{can} \quad (9)$$

These equations,  $n^{heal}$  and  $n^{can}$  denote the healthy and cancerous areas of the output image  $n$ .  $\sigma_{1,2}$  denote the characteristic map representations obtained from the fully convolutional layers. The appearance of the healthy area is improved using Eq. (10) as it occupies most of the part compared to the cancerous area.

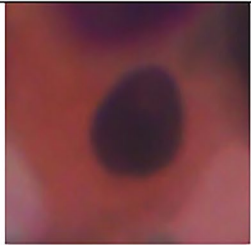

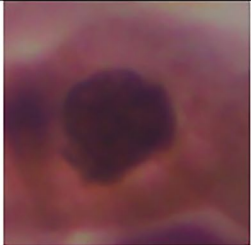
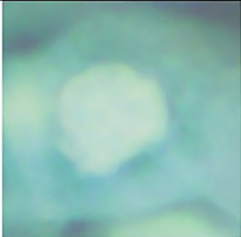
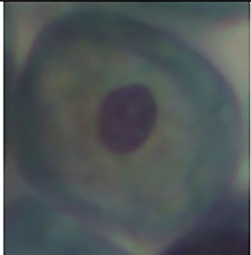
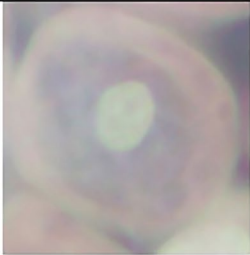
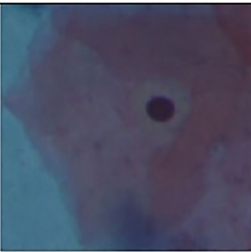

$$L_{ai} = H [||n^{heal} - m_j^{heal}||] \quad (10)$$

The discriminator function is employed in this model for producing more appealing imagery of the cervical regions and its corresponding loss function is formulated as an aggregation of globalized as well as localized loss functions as given in Eqs. (11)–(13),

$$L_{dl} = L_{gb} + L_{lc} \quad (11)$$

$$L_{gb} = H [(G_d(m_j, l_j)) - 1]^2 + H [(G_d(n, l_{ij}))^2] \quad (12)$$



SAMPLE INPUT		PRE-PROCESSED IMAGE	
			
			
			
			

**Fig. 2.** Pre-processing results.

$$L_{lc} = H \left[ (L_{lc} (m_j^{loc})) - 1 \right]^2 + H \left[ (L_{lc} (n^{loc}))^2 \right] \quad (13)$$

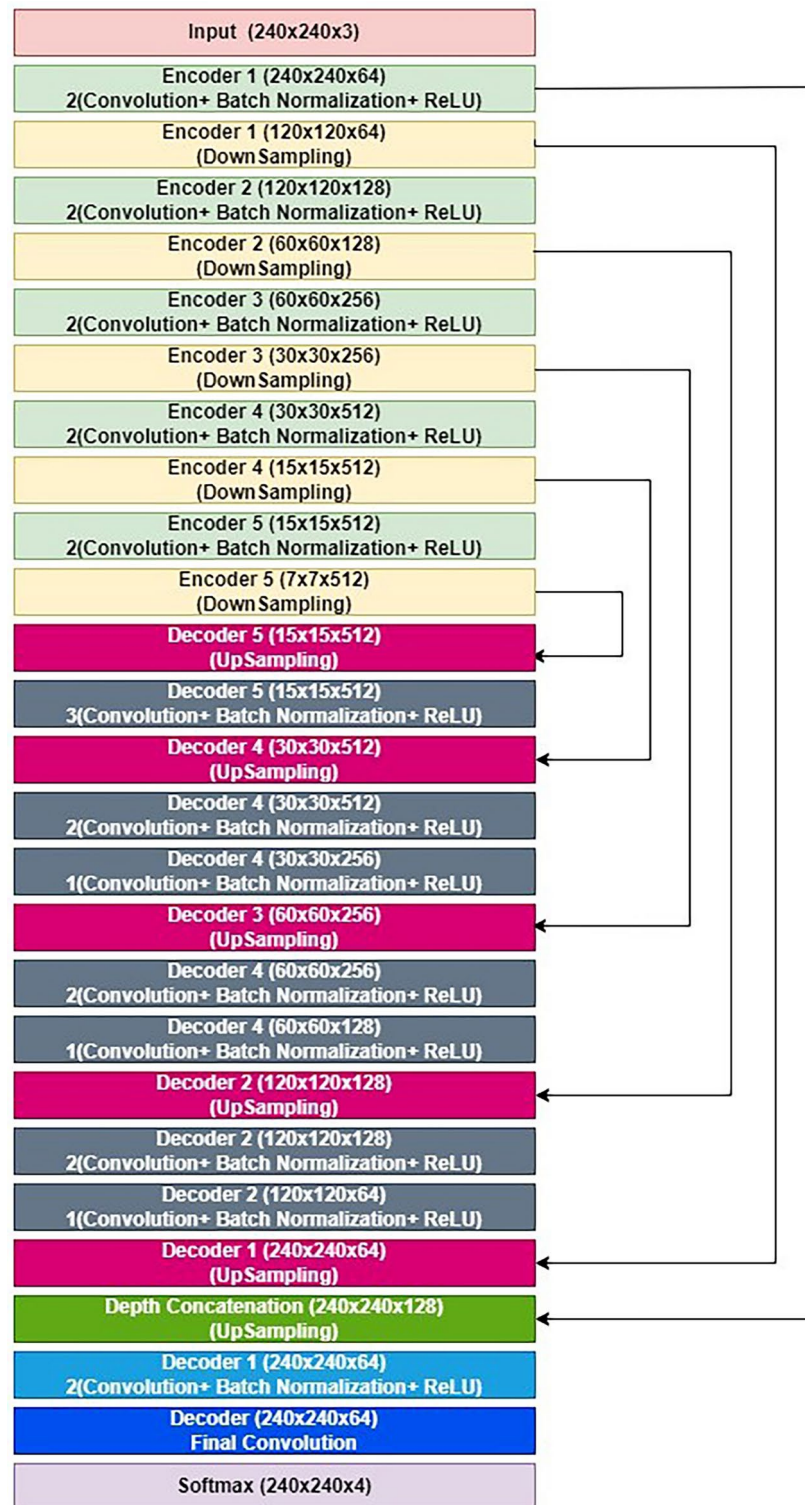
The aggregate loss function of the m-GAN model is given in Eq. (14),

$$L_{m-GAN} = \alpha L_{da} + \beta L_{ai} + \gamma L_{dl} \quad (14)$$

The steps involved in the m-GAN model for data augmentation are summarized in Fig. 3.

### Segmentation phase

The contrast-enhanced—augmented images are fed into the Segmentation phase. In this stage, the segmentation process of cancerous regions from a cervical cavity is carried out using the Seg-UNet<sup>20</sup> model. The unique SegNet and U-Net structures commonly employed for visual segmentation are combined to create the Seg-UNet model. An encoding module for decreasing the sample size and a decoding function for increasing the sample size make up the SegNet model. Although this model includes an additional convolutional layer in each kernel block, its decoding and encoding algorithms are comparable to the thirteen convolutional layers of the VGG16 model. It is provided to correspond with the characteristic map area of the fourth dimension in the convolution layer. A delay link is added to the initial kernel structures of the SegNet model, a modification inspired by U-Net<sup>21</sup>. The architecture of the Seg-UNet model is presented in Fig. 4, and the outcomes obtained from this phase are shown in Fig. 5.



**Fig. 3.** m-GAN model algorithm.

### Feature extraction phase

The denoising autoencoder is employed in this work to extract the features from the segmented images. Initially, the segmented input ( $S$ ) is converted into a falsified input ( $\tilde{s}$ ) by introducing the arbitrary delineation as  $\tilde{s} \sim r_T(\tilde{s}|s)$ . In the present investigation, obscuring distortion is instigated, meaning that the numerical value of every component within an unaltered source sample  $S$  is arbitrarily restricted to zero by a predetermined proportion. It might be conceptualised as either replacing those absent elements or establishing the standard value for addressing values that are not present. Consequently, the specific entry channel loses all information

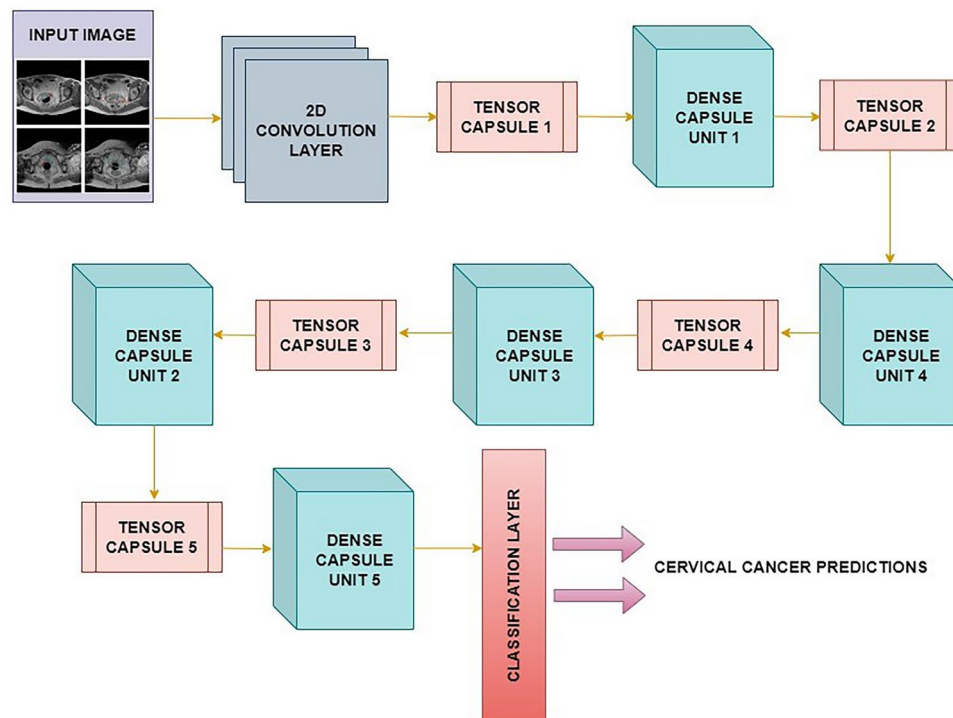


Fig. 4. Seg-UNet architecture.

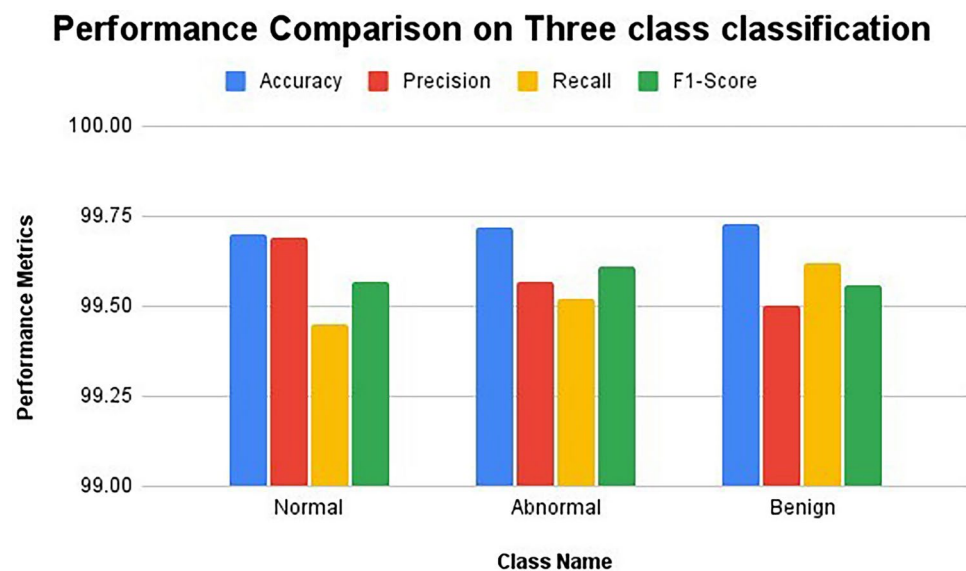


Fig. 5. Segmentation results.

about these concealing elements, and the denoising autoencoder can be considered as a tool provided to fill in these individually generated empty spaces. Moreover, when the value of a node is set to zero, it indicates that it will not be taken into account at all when determining the number of neurons in the subsequent layer. The initialization function for the autoencoder is computed using the formulation given in (15),

$$p = g_o(\tilde{s}) = h(w_{t1}\tilde{s} + b_{t1}) \quad (15)$$

The function  $h$  used in the Eq. (15) can be mathematically represented as shown in Eq. (16),

$$h(s) = \frac{1}{1 + e^{-s}} \quad (16)$$



When  $p$  is fed as input to the decoding function, the output  $q$  is determined using the Eq. (17) as,

$$q = r_0(p) = h(w_{t2}p + b_{t2}) \tag{17}$$

$q$  in the above equation can be considered as the recreated output for the input  $\tilde{s}$ . The loss function for identifying the difference between the original and falsified input is computed using the Eq. (18),

$$L(s, q) = s \log(q) + (1 - s) \log(1 - q) \tag{18}$$

The Mini-batch Gradient Descent method is used to maximise the variables after they are arbitrarily created. The loss due to recreation between  $s$  and the rebuilt  $q$  from  $p$  is mitigated using the denoising autoencoder and thus facilitates the effective extraction of features. The hyperparameter values are listed in Table 2.

Classification phase

Dense CapsNet model is used in this work to classify the Cervical cancer imagery. This model comprises of a two-dimensional convolutional layer along with the tensors in a layer. These tensors are grouped together with dense connectivity to form the Dense Capsule layers. The architecture of the Dense CapsNet model is presented in Fig. 6. In this model, the input images are fed to the convolutional layer in order to create the characteristic maps. Each of these characteristic maps are represented in the form of tensors which take the structure in the form of (length, breadth, weight) denoted as (l, b, w). There are five tensor capsule layers present in the Dense CapsNet model. Each of the tensors with the (l, b, w) structure is passed on to the first tensor layer. Further, it is moved on to the Dense Capsule layer. Any unit  $k$  in the dense capsule layer receives  $T_{k-1}$  tensors which take the form  $(l^{k-1}, b^{k-1}, w^{k-1})$  and the corresponding outcomes are of the form  $(l^k, b^k, w^k)$  by taking multiple algorithm executions. The topological proportions of the tensors are minimized by including descends of the size (3,3). The output from the fifth tensor layer and dense capsule layer is the input to the final categorization layer in the model. The classifications from this layer are in the form of  $(o^{num}, s^{num})$ , where  $o^{num}$  denotes the count of the classes to be categorized by the model and  $s^{num}$  denotes the identification number associated with each of the outputs.

Results and discussion

This section presents the results obtained on evaluating the proposed approach and applying it to the SIPaKMeD dataset.

Dataset description

The SIPaKMeD dataset consists of images of the Pap Smear slides of the cervical cavity. According to the visual impression of the images and cellular structures, these cervical cells present in the dataset are classified into five different classes, as shown in Table 3. Based on these categories, the cells are grouped as normal, abnormal, or benign. The dataset comprises an aggregate of 4049 images of the cellular forms of cervical regions. The entire dataset is divided for training and testing purposes in a ratio of 80:20. The data augmentation methods employed in this research increased the dataset utilized for training purposes by 5. This dataset can be accessed using the following link, <https://www.cs.uoi.gr/marina/sipakmed.html>.

Phase	Hyperparameter	Value/Range
Preprocessing phase	Pixel expansion factor ( $\delta$ )	20
	Threshold values ( $\theta$ )	$G1 \leq 99, 100 \leq G2 \leq 169, G3 \geq 170$
	Contrast limits (H_z)	Based on histogram normalization
Data augmentation phase	GAN loss weights ( $\alpha, \beta, \gamma$ )	0.3, 0.4, 0.3 (experimentally tuned)
	Learning rate (GAN)	0.0002
Segmentation phase	Kernel size	$3 \times 3$
	Number of filters	{32, 64, 128}
	Batch size	16
	Learning rate	0.001
Feature extraction phase	Denoising ratio	0.3 (30% data masked)
	Encoding/Decoding units	512 (encoder), 512 (decoder)
Classification phase	Capsule dimensions (l, b, w)	(64, 64, 128)
	Number of tensor layers	5
	Batch size	32
	Learning rate	0.0001
	Dropout rate	0.3

Table 2. Hyperparameter values of the proposed model.

**Algorithm of m-GAN model**  
Input: Cervical dataset  
Output: Augmented images  
Step 1: Load the dataset  
Step 2: Choose the  $k^{\text{th}}$  portion of image  $m_i$   
Step 3: Retrieve the attribute label from the cancerous region of  $m_i$  as  $l_c$   
Step 4: Choose the  $k^{\text{th}}$  portion of image  $m_j$   
Step 5: Retrieve the attribute label from the structural region of  $m_j$  as  $l_s$   
Step 6: Concatenate  $l_{ij} = l_c + l_s$   
Step 7: Input  $l_{ij}$  and  $m_i$  to the m-GAN *generator* to obtain the output image  $n$   
Step 8: Compute Area based loss  $L_{da}$  and  $L_{ai}$  using equations (7) to (10)  
Step 9: Input  $n$  and  $m_j$  to m-GAN *localized discriminator* and compute localized loss  $L_{lc}$   
Step 10: Input  $(m_j, l_j)$  and  $(n, l_{ij})$  to m-GAN *globalized discriminator* and compute globalized loss  $L_{gb}$   
Step 11: return augmented images

Fig. 6. Dense CapsNet model.

Cellular types	Cellular categories	Cellular image count
Koilocytotic	Abnormal	825
Metaplastic	Benign	793
Dyskeratotic	Abnormal	813
Parabasal	Normal	787
Superficial	Normal	831

Table 3. Dataset details.

Experimental evaluation

The appropriate verification method was crucial for the dataset containing 4049 cellular images. Using a holdout verification procedure, 20% of the data was retained for testing and 80% for training requirements. The holdout verification procedure is the most widely used strategy that yields satisfactory outcomes. To help the model train more quickly, the holdout technique usually entails dividing the dataset into an initial set for training and a set to be evaluated. The Dense CapsNet model was trained on most of the dataset, and its performance was assessed on the remaining part. Using the holdout approach, the model was trained for several epochs, and it was evident that there was a steady increase in accuracy for training and validation datasets; subsequently, loss was minimized in consecutive epochs. Figure 7 shows the accuracy and loss of the model.

The confusion matrix that depicts the predictions made by the Dense CapsNet model in test set model for three class classification is given in Fig. 8. The model classified 1613 images as normal correctly out of the total 1618 images that represented normal class and there were five misclassifications concerning normal class. 4 normal images were classified to be abnormal and 1 was misclassified as benign. The model classified the 1638 abnormal images in the dataset as 1631 to be abnormal, 5 to be normal, and 2 to be benign. 4 benign images were misclassified as normal and the remaining 789 benign images were classified correctly.

The performance of the model was evaluated using performance metrics such as Matthews correlation coefficient(MCC), accuracy, precision, recall and F1 score. The outcomes obtained for the Three Class Classification are presented in Table 4 and depicted graphically in Fig. 9. An accuracy of 99.69% was obtained was normal classifications with precision of 99.62%, recall of 99.45%, and F1-score of 99.57%. The abnormal classes were detected with accuracy, precision, recall and F1 score of 99.57%, 99.47%, 99.52%, and 99.61% respectively. The Benign class was identified 99.5% accurately with 99.43% precision, 99.62% recall and 99.56% F1 score.

Furthermore, performance of the Dense CapsNet model for the five-class classification in test set was also examined and the obtained confusion matrix is presented in Fig. 10.

There were 823 correct classifications for Koilocytotic class with one misclassification as metaplastic and parabasal each. There was a total of 793 images for Metaplastic class in the dataset and 792 were classified appropriately as Metaplastic and one sample was classified inaccurately as parabasal. Four misclassifications

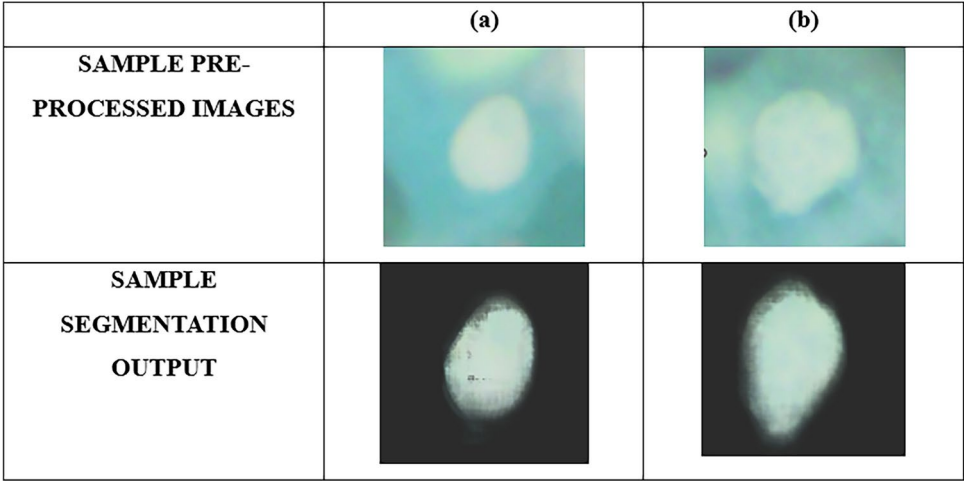


Fig. 7. Accuracy Vs Loss.

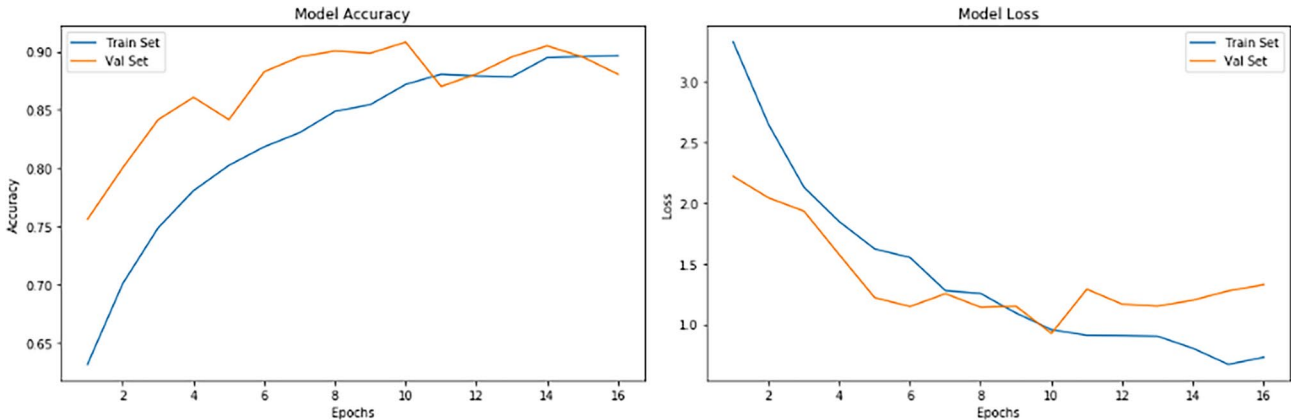


Fig. 8. Confusion matrix for three class classification.

Class name	Accuracy	Precision	Recall	F1-Score	MCC
Normal	99.69	99.62	99.45	99.57	0.9923
Abnormal	99.57	99.47	99.52	99.61	0.9919
Benign	99.5	99.43	99.62	99.56	0.9916
Average	99.60	99.56	99.53	99.58	0.9920

Table 4. Performance metrics for three class classification.

were obtained for Dyskeratotic class and the remaining 809 samples were predicted precisely. 782 samples of Parabasal class and 829 samples of Superficial class were detected correctly with higher accuracy and a smaller number of incorrect predictions.

The results obtained by each of the class in Five class classifications are presented in Table 5 and Fig. 11. It can be noted that the Dense CapsNet model exhibited an accuracy of 99.76% in predicting Koilocytotic class, 99.87% for Metaplastic class, 99.51% for Dyskeratotic class, 99.36% for Parabasal class and 99.76% for Superficial class. Similarly, the precision, recall and F1 scores were also found to be higher for all the five classes as the number of misclassifications made by the Dense CapsNet model is minimal. The average accuracy, precision, recall and F1 score achieved by the model are 99.65%, 99.58%, 99.62% and 99.67% correspondingly.

Additionally, the performance of the model is also compared against the existing research in the Table 6. The authors in<sup>24</sup> used the ANFIS classifier to detect cervical cancer and achieved an accuracy of 93.2% with F1 score of 92.8%. The researchers in<sup>26</sup> produced an accuracy of 94.5% for the fuzzy ensemble model with 93.6% precision, 93.8% recall, and 94.2% F1 score. The deep model based on feature fusion proposed in<sup>27</sup> obtained 95.6% accurate results for the detection of tumors in a cervical cavity. The model employed in<sup>28</sup> was based

Testing set				
TARGET \ OUTPUT	Normal	Abnormal	Benign	SUM
Normal	1613 39.84%	4 0.10%	1 0.02%	1618 99.69% 0.31%
Abnormal	5 0.12%	1631 40.28%	2 0.05%	1638 99.57% 0.43%
Benign	4 0.10%	0 0.00%	789 19.49%	793 99.50% 0.50%
SUM	1622 99.45% 0.55%	1635 99.76% 0.24%	792 99.62% 0.38%	4033 / 4049 99.60% 0.40%

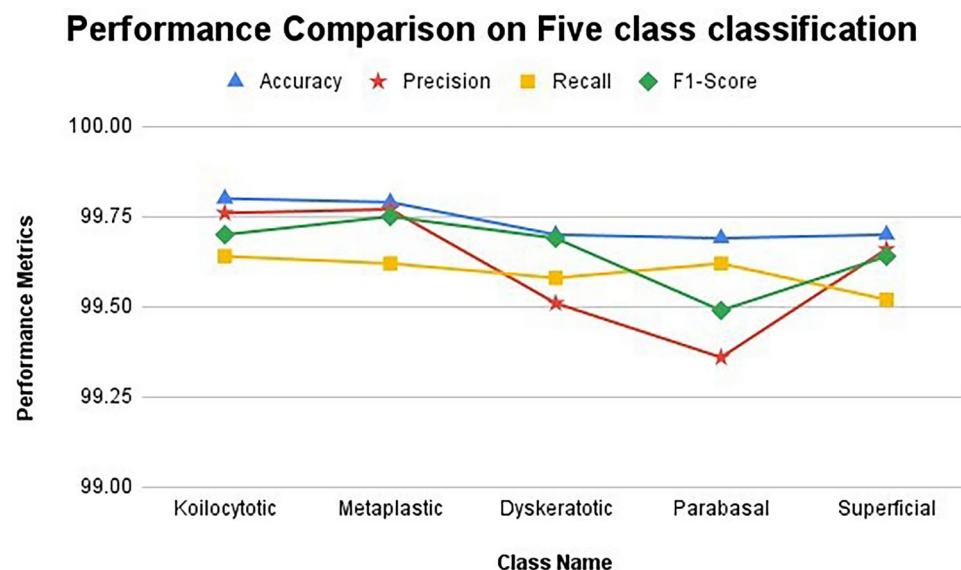
Fig. 9. Performance comparison of three class classification.

Testing set						
TARGET \ OUTPUT	Koilocytotic	Metaplastic	Dyskeratotic	Parabasal	Superficial	SUM
Koilocytotic	823 20.33%	1 0.02%	0 0.00%	1 0.02%	0 0.00%	825 99.76% 0.24%
Metaplastic	0 0.00%	792 19.56%	0 0.00%	1 0.02%	0 0.00%	793 99.87% 0.13%
Dyskeratotic	1 0.02%	0 0.00%	809 19.98%	1 0.02%	2 0.05%	813 99.51% 0.49%
Parabasal	1 0.02%	1 0.02%	1 0.02%	782 19.31%	2 0.05%	787 99.36% 0.64%
Superficial	1 0.02%	1 0.02%	0 0.00%	0 0.00%	829 20.47%	831 99.76% 0.24%
SUM	826 99.64% 0.36%	795 99.62% 0.38%	810 99.88% 0.12%	785 99.62% 0.38%	833 99.52% 0.48%	4035 / 4049 99.65% 0.35%

Fig. 10. Confusion matrix for five class classification.

Class name	Accuracy	Precision	Recall	F1-Score
Koilocytotic	99.76	99.66	99.64	99.7
Metaplastic	99.87	99.57	99.62	99.75
Dyskeratotic	99.51	99.42	99.58	99.69
Parabasal	99.36	99.26	99.62	99.49
Superficial	99.76	99.66	99.52	99.64
Average	99.65	99.58	99.62	99.67

**Table 5.** Performance metrics for five class classification.



**Fig. 11.** Performance comparison of five class classification.

References	Accuracy	Precision	Recall	F1-Score
Jaya et al. <sup>24</sup>	93.2	–	–	92.8
Manna et al. <sup>26</sup>	94.5	93.6	93.8	94.2
Rahaman et al. <sup>27</sup>	95.6	94.1	94.3	95.3
Jiang et al. <sup>28</sup>	96.9	–	–	96.5
Yu et al. <sup>30</sup>	97.6	96.2	96.4	97.1
Proposed	99.65	99.58	99.62	99.67

**Table 6.** Comparison of existing vs proposed approaches.

on fully convolutional neural networks that achieved an accuracy of 96.9%. The investigators in<sup>30</sup> proposed a hybrid deep learning framework for the identification and categorization of cervical cancer with 97.6% accuracy. However, it was evident from experimental evaluation and comparison with existing methods that the proposed model produced superior performance outcomes than the other models in the literature.

### Limitations of present research

Compared to CNNs, Dense CapsNet models are typically compute-intensive and cost-prohibitive, especially during training. Iterative procedures are used in the dynamic routing method of Dense CapsNet model, which increases information processing overhead. This degree of abstraction can be detrimental for real-time applications or situations with limited resources. Hyperparameters like the quantity of capsules, routing variations, and retention rates have an impact on dense capsule networks. Comprehensive testing and computational capabilities are frequently needed to determine the ideal collection of hyperparameters, which may not be possible in many situations.



## Conclusion

Cervical cancer must be detected promptly to reduce the number of women who suffer deaths from this often-recurring kind of cancer. Computerized systems are frequently used to help physicians identify this kind of cancer in its infancy. This work successfully diagnosed cervical cancer using the SIPaKMeD dataset, which includes imagery of the cervix collected from pelvic smear tests, by utilizing well-known deep learning techniques. The suggested work uses the m-GAN technique for data augmentation, Seg-UNet for segmentation, denoising autoencoders for feature extraction, and the Dense CapsNet model for cervical cancer classification. The performance of the proposed approach is evaluated for multi-class classifications such as 3-class classification and 5-class classification. The average accuracy obtained for 3-class and 5-class classifications is 99.6% and 99.65%, respectively. As an extension to the present research, Artificial intelligence driven techniques can be employed to detect and categorize cervical cancer in order to minimize the restrictions related to data processing and computational resources.

## Data availability

The datasets generated during and/or analysed during the current study are available in the [kaggle] repository, [<https://www.cs.uoi.gr/marina/sipakmed.html>].

Received: 13 October 2024; Accepted: 4 December 2024

Published online: 30 December 2024

## References

- Jhingran, A. et al. *Cancers of the Cervix, Vulva, and Vagina, in Abeloff's Clinical Oncology* 1468–1507 (Elsevier, 2020).
- Alyafei, Z. & Ghouti, L. A fully-automated deep learning pipeline for cervical cancer classification. *Exp. Syst. Appl.* **141**, 112951. <https://doi.org/10.1016/j.eswa.2019.112951> (2020).
- Zou, Y. et al. Precision matters: the value of PET/CT and PET/MRI in the clinical management of cervical cancer. *Strahlenther. Onkol.* (2024).
- Cheng, Z. et al. Application of serum SERS technology based on thermally annealed silver nanoparticle composite substrate in breast cancer. *Photodiagn. Photodyn. Therapy* **41**, 103284 (2023).
- Su, Y. et al. Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis. *Comput. Biol. Med.* **145**, 105409 (2022).
- Ma, X. et al. Detection of breast cancer based on novel porous silicon Bragg reflector surface-enhanced Raman spectroscopy-active structure. *Chin. Opt. Lett.* **18**, 5 (2020).
- Chikhaoui, M., Smail, F., Aissa, A. B. S., Benhamida, H. & Hamri, R. A retrospective study on cervical cancer screening using pap smear and related factors among women living in tiaret, Algeria. *Indian J. Gynecol. Oncol.* **18**(4), 1–8 (2020).
- Abdullah, A. A., Fonetta, A., Giong, D., Adilah, N. & Zahri, H. Cervical cancer detection method using an improved cellular neural network (CNN) algorithm. *Indonesian J. Elect. Eng. Comput. Sci.* **14**(1), 210–218. <https://doi.org/10.11591/ijeecs.v14.i1.pp210-218> (2019).
- Li, X. et al. A comprehensive review of computer-aided whole-slide image analysis: From datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* **55**, 4809–4878 (2022).
- Gharsellaoui, S., Mansouri, M., Refaat, S. S., Abu-Rub, H. & Messaoud, H. Multivariate features extraction and effective decision making using machine learning approaches. *Energies* **13**(3), 609 (2020).
- Miyagi, Y., Takehara, K., Nagayasu, Y. & Miyake, T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncol. Lett.* **2019**, 1602–1610. <https://doi.org/10.3892/ol.2019.11214> (2019).
- Ples, L. et al. The accuracy of cytology, colposcopy and pathology in evaluating precancerous cervical lesions. *Diagnostics* **12**(8), 1947. <https://doi.org/10.3390/diagnostics12081947> (2022).
- Andijany, A. A. et al. Prevalence of abnormal pap smears in the western region of Saudi Arabia from 2010 to 2022. *Cytojournal* **20**(44), 2023 (2023).
- Hu, L. et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J. Natl. Cancer Inst.* **111**(9), 923–932. <https://doi.org/10.1093/jnci/djy225> (2019).
- Gaona, Y. J. et al. Radiomics diagnostic tool based on deep learning for colposcopy image classification. *Diagnostics* **12**(7), 1694. <https://doi.org/10.3390/diagnostics12071694> (2022).
- Bai, B. et al. Biomedical signal processing and control detection of cervical lesion region from colposcopic images based on feature reselection. *Biomed. Signal Process. Control* **57**, 101785. <https://doi.org/10.1016/j.bspc.2019.101785> (2020).
- Devi, M. A., Sheeba, J. I. & Joseph, K. S. 'Neutrosophic graph cut-based segmentation scheme for efficient cervical cancer detection. *J. King Saud Univ.-Comput. Inf. Sci.* **34**(1), 1352–1360. <https://doi.org/10.1016/j.jksuci.2018.09.014> (2022).
- Zhang, T. et al. Biomedical signal processing and control cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images. *Biomed. Signal Process. Control* **55**, 101566. <https://doi.org/10.1016/j.bspc.2019.101566> (2020).
- Skerrett, E. et al. 'Multicontrast pocket colposcopy cervical cancer diagnostic algorithm for referral populations'. *BME Front.* **2022**, 1–13. <https://doi.org/10.34133/2022/9823184> (2022).
- Bing, P. et al. A novel approach for denoising electrocardiogram signals to detect cardiovascular diseases using an efficient hybrid scheme. *Front. Cardiovasc. Med.* **11**, 456 (2024).
- Zhou, J. et al. Network pharmacology combined with experimental verification to explore the potential mechanism of naringenin in the treatment of cervical cancer. *Sci. Rep.* **14**, 1 (2024).
- Elayaraja, P. & Suganthi, M. Automatic approach for cervical cancer detection and segmentation using neural network classifier. *Asian Pac. J. Cancer Prevent.* **19**(12), 3571–3580. <https://doi.org/10.31557/APJCP.2018.19.12.3571> (2018).
- Chen, S., Chen, Y., Yu, L., & Hu, X. YTHDC1 inhibits cell proliferation and angiogenesis in cervical cancer by regulating m6A modification of SOCS4 mRNA. *Mol. Cell. Toxicol.* **20**(3), 533–540 (2024).
- Jaya, B. K. & Kumar, S. S. 'Image registration based cervical cancer detection and segmentation using ANFIS classifier'. *Asian Pac. J. Cancer Prevent.* **19**(11), 3203–3209. <https://doi.org/10.31557/APJCP.2018.19.11.3203> (2018).
- Tan, X. et al. Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study. *Cancer Cell Int.* **21**(1), 1–10. <https://doi.org/10.1186/s12935-020-01742-6> (2021).
- Manna, A., Kundu, R., Kaplun, D., Sinitca, A. & Sarkar, R. A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. *Sci. Rep.* **11**(1), 14538. <https://doi.org/10.1038/s41598-021-93783-8> (2021).
- Rahaman, M. M. et al. DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Comput. Biol. Med.* **136**, 104649. <https://doi.org/10.1016/j.combiomed.2021.104649> (2021).

28. Jiang, X. et al. MRI based radiomics approach with deep learning for prediction of vessel invasion in early-stage cervical cancer. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **18**(3), 995–1002 (2021).
29. Luan, S. et al. Deep learning for fast super-resolution ultrasound microvessel imaging. *Phys. Med. Biol.* **68**(24), 245023 (2023).
30. Yu, X. et al. Deep learning for fast denoising filtering in ultrasound localization microscopy. *Phys. Med. Biol.* **68**(20), 205002 (2023).
31. Zhou, L. et al. Multi-omics graph convolutional networks for digestive system tumour classification and early-late stage diagnosis. *CAAI Trans. Intell. Technol.* (2024).
32. Fatima, M. et al. Breast lesion segmentation and classification using U-Net saliency estimation and explainable residual convolutional neural network. *Fractals*. <https://doi.org/10.1142/S0218348X24400607> (2024).
33. Rauf, F. et al. Artificial intelligence assisted common maternal fetal planes prediction from ultrasound images based on information fusion of customized convolutional neural networks. *Front. Med.* **11**, 1486995 (2024).
34. Ullah, M. S., Khan, M. A., Albarakati, H. M., Damaševičius, R. & Alsenan, S. Multimodal brain tumor segmentation and classification from MRI scans based on optimized DeepLabV3+ and interpreted networks information fusion empowered with explainable AI. *Comput. Biol. Med.* **182**, 109183 (2024).
35. Li, Y. et al. AI-based visual speech recognition towards realistic avatars and lip-reading applications in the metaverse. *Appl. Soft Comput.* **164**, 111906 (2024).

## Acknowledgements

This study is supported via funding from Prince Sattam Bin Abdulaziz University project number (PSAU/2024/R/1446). Nisreen Innab would like to express sincere gratitude to AlMaarefa University, Riyadh, Saudi Arabia, for supporting this research.

## Author contributions

Hui Yang: Conceptualization, Methodology, Formal analysis, Validation, Resources, Supervision, Writing—original draft, Writing—review & editing. Walid Aydi: Validation, Resources, Supervision, Writing—original draft, Writing—review & editing. Nisreen Innab: Formal analysis, Validation, Resources, Supervision, Writing—original draft, Writing—review & editing. Mohamed E. Ghoneim: Validation, Resources, Writing—original draft, Writing—review & editing. Massimiliano Ferrara: Validation, Resources, Writing—original draft, Writing—review & editing.

## Funding

This research received no specific grant from any funding agency.

## Competing interests

The authors declare that they have no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to H.Y. or M.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024