

Deep Learning for Cervical Cancer Detection

Posted on June 8, 2026

ABSTRACT

Cervical cancer (CC) is considered as a second leading cause of cancer death in women.

Though it seems that this cancer only affects elderly women, it actually affects younger women also. Since it has four stages, the Early detection may result in less mortality rates. There exist various methods for the detection of such diseases. To overcome the drawbacks in the methods in use and to diagnose this disease precisely, a deep learning approach called cervical cancer detection via CNN-trans-attention layer net is recommended for the accurate diagnosis of the disease. Initially the input data is pre-processed by which the noises are removed and the model's robustness is improved followed by feature extraction by which the shape, texture, and color-based features are extracted from the pre-processed images. Finally, classification is done by combining Convolutional Neural Networks (CNNs) with Transformer-based self-attention mechanisms, enabling it to effectively capture intricate patterns in cervical images. The suggested approach has an accuracy of 98.7%. Thus, from the results, it is seen that our suggested approach outperforms well in comparison to the other existing methods.

1. INTRODUCTION

Cervical Cancer is a disease that affects more than half a million people annually and causes more deaths globally. The female reproductive system, specifically the cervix (uterus), is the focal point of CC. Though it seems that this cancer only affects elderly women, it actually affects women as young as 20 to 30 years old [1]. The second most frequent malignancy and the second leading cause of cancer death in women is CC [2]. Human papillomavirus (HPV) infections are the main cause of almost all cases of CC. The most prevalent STD in the world, HPV is the main factor in almost all occurrences of CC. Both men and women should obtain HPV vaccinations to prevent the development of CC, which is an efficient primary prevention method against the disease, although vaccination rates are still low [3]. Depending on spreading level CC has four stages: In Stage 1 CC only spreads to the lymph nodes; stage 2 larger cancer spreads outside of the uterus and cervix or to the lymph nodes; stage 3 CC spreads to the lower part of the vagina or the pelvis along with blocking the ureters, the tubes that carry urine from the kidneys to the bladder; and stage 4 CC spreads outside of the pelvis to organs such as your lungs, bones, or liver [4]. A greater proportion of CC deaths is caused by the fact that 90% of cervical malignancies occur in low- and middle-income nations, where there are no organised screening or HPV vaccination initiatives [5]. Therefore, a key method for completely controlling CC is early diagnosis. CC is discovered via a variety of diagnostic tests. Among them, colposcopy, biopsy,

and cone biopsy are all highly efficient. The primary screening procedure for CC is the Pap test [6].

The obstacles preventing the wider use of our current systems should be addressed by the diagnostic techniques deployed. 80-95% of women with early-stage disease (stages I and II) and 60% of women with stage III disease can be cured with surgery or chemoradiotherapy [7]. Medical image analysis, which involves taking photographs of the bodily structures and processes, is done for diagnostic purposes. When diagnosing a patient's ailments, machine learning for medical picture analysis offers a number of advantages [8]. A test called magnetic resonance imaging (MRI) gives 2D or 3D pictures of the inside of the body. The most cutting-edge medical imaging method uses radio waves and a strong magnetic field to produce high-quality images of each affected body part by enabling visualisation of the finer features of interior structure. When treating CC or other malignancies, MRI is frequently employed [9]. For the purpose of detecting CC, machine learning algorithms can be applied at two separate stages. The first instance is when it is employed after meaningful feature extraction. Machine learning algorithms can be employed in this situation to learn these qualities, which will help forecast and diagnose the malignancy. The second scenario is when machine learning techniques are utilised to extract crucial features, with feature learning followed by CC prediction and detection [10]. On these altered datasets, various classifiers were used, and the ML techniques performed well. Additionally, we investigated how data transformation might improve classifier performance. Then, different feature selection techniques (FST) were applied to these altered datasets to identify classifiers that performed optimally when taking CC risk variables into account [11]. The subset of AI and ML known as deep learning (DL) consists of numerous layers of computational models for data processing that facilitate learning by presenting input data at different levels of abstraction. Recently, DL has been effectively applied to address real-world problems in a wide range of applications. DL approaches are a great way to advance medical image analysis in a variety of scientific and clinical settings, like the identification and treatment of various tumours [12]. Methods based on DL have gained importance for categorising CC patients into various risk groups. In order to make precise decisions based on predictive models, many techniques, including ANN, Bayesian networks, SVM, and decision trees, have been used in cancer research for a long time [13].

Furthermore, 85% of CC cases were identified at an advanced stage. The main causes of this disease are the restricted availability of health services and the general lack of knowledge about CC. Therefore, early detection of CC is the foundation for successful therapies, but doing so requires efficient methods. The key factors used to lower mortality and morbidity for efficient screening are the standard of the screening test, the choice of lesion treatment, and access to facilities [14]. In conclusion, CC continues to be the fourth most prevalent malignancy in women worldwide and a major cause of morbidity and mortality in the United States. The most vulnerable groups continue to have unequal rates of immunisation, screening, and treatment, which leads to inferior results. These groups include racial and ethnic minorities, the socioeconomically disadvantaged, and people who live in rural or remote locations. The WHO has developed a plan to address each of these issues in order to eradicate CC by 2030. We will get closer to this objective through improving education, access to healthcare, and the growth of screening and immunisation programmes locally and worldwide [15]. The contribution of the work is as follows,

- Using the dataset, pre-processing, segmentation, feature extraction and classification is done.
- Input data is pre-processed for the removal of noise and to improve model's robustness, followed by feature extraction by which the shape, texture, and color-based features are extracted from the pre-processed images. Finally, classification is done by combining Convolutional Neural Networks (CNNs) with Transformer-based self-attention mechanisms, enabling it to effectively capture intricate patterns in cervical images.
- The recommended method and existing methods are compared and the results are obtained based on the evaluation metrics.

The remaining section of the paper is as follows. The Section 2 explains about the Related Work, Section 3 explains the proposed part and Section 4 shows the Results and Discussions and finally Section 5 gives the Conclusion of the paper.

2. REALTED WORK

R. Kavitha et.al [16] explains that a screening process called as a Pap test is frequently carried out in order to find CC in women in its earliest stages. This article explains how to use Brightness Preserving Dynamic Fuzzy Histogram Equalisation to enhance photos. The fuzzy c-means technique is used to identify individual components and identify the appropriate region of interest. To identify the proper area of interest, the photos are segmented using the fuzzy c-means technique. The ACO algorithm serves as the feature selection algorithm, and as a result, the suggested method performs better and achieves a greater accuracy rate.

Zahid Hasan Ontor et.al [17] proposed a DL -based intelligent method to use real-time photos to detect CC in its early stages. CC Pap-smear test image datasets were acquired, labelled, and pre-processed in order to develop the model. After that, the labelled dataset was used to train the YOLOv5 model. The most effective model for creating an intelligent system to detect CC at an early stage was determined in this study by applying three of the most recent iterations of the YOLOv5 model. YOLOv5s surpassed all other applied models with a precision and recall value of 0.8279 and 0.8265, respectively. The results of the study show that the proposed approach has a great deal of potential for early-stage CC diagnosis utilising real-time imagery. Mavra Mehmood et.al [18] introduced a concept called CervDetect uses machine learning algorithms to assess the risk factors for malignant cervical development. The data is pre-processed by CervDetect using Pearson correlation between the input variables and the output variable. CervDetect chooses important features using the random forest (RF) feature selection method. In order to identify CC, CervDetect employs a hybrid strategy that combines RF and shallow neural networks. CervDetect beats state-of-the-art studies in its ability to predict CC, as evidenced by its accuracy of 93.6%, MSE of 0.07111, FPR of 6.4%, and FNR of 100%.

K. M. A. Adweb et.al [19] To perform CC screening, very deep residual learning-based networks have been proposed. Additionally, we emphasise the significance of the activation functions on the

effectiveness of a residual network (ResNet) in this work. As a result, three residual networks with the same structure and various activation functions are constructed. Using a dataset of colposcopy cervical images, the employed models were trained and tested. The experimental results revealed that designed residual networks with leaky and parametric rectified linear unit (Leaky-RELU and PReLU) activation functions performed nearly equally well in terms of accuracy, reaching accuracies of 90.2 and 100%, respectively. When this high accuracy was compared to the results of other studies in the field, it demonstrated superior performance in the detection of precancerous and healthy colposcopy cervical pictures.

R Elakkiya et.al [20] paper, discusses an effective hybrid DL technique using Small-Object Detection-Generative Adversarial Networks (SOD-GAN) with Fine-tuned Stacked Autoencoder (F-SAE) is developed to address the shortcomings mentioned above. The SOD-GAN's generator and discriminator are created using RCNNs, or region-based convolutional neural networks. To speed up lesion detection, the SOD-GAN hyperparameters are normalised and optimised while the model parameters are adjusted using F-SAE. Without any aid from initial classification and segmentation, the suggested approach automatically identifies and categorises cervical premalignant and malignant diseases based on deep features. Multivariate heterogeneous data have also been the subject of much testing, and the proposed method has demonstrated potential improvements in efficiency and decreases in time complexity.

Yue Ming et.al [21] proposed a computer-aided DL-based framework to detect CC using multimodal medical images to increase the efficiency of clinical diagnosis. Image registration, multimodal image fusion, and lesion object identification make up this framework's three parts. Our adaptive image fusion technology merges multimodal medical pictures differently than conventional methods. state-of-the-art (SOTA) object-detection DL-based methods in images with diverse modalities are done, and we conduct comprehensive experiments to compare the performance of other image fusion approaches with these SOTA methods. Our suggested strategy improves the recognition accuracy of multiple item detection models by an average of 6.06% when compared to PET, which has the highest recognition accuracy in single-modality images. And when compared to the best outcomes of existing multimodal fusion techniques, our outcomes show an improvement of 8.9% on average. Yao Xiang et.al [22] This paper presents an efficient and totally segmentation-free method for automated cervical cell screening that utilizes modern object detector to directly detect cervical cells or clumps, without the design of specific hand-crafted feature. We cascade an additional task-specific classifier to enhance the classification performance of hard examples, which are four extremely similar categories. On cervical cell image-level screening, our model achieves 97.5% sensitivity (Sens) and 67.8% specificity (Spec). Additionally, we increase the average precision (AP) of hard cases, which are the most valuable but hardest to differentiate, and achieve a best mean average precision (mAP) of 63.4% on cervical cell-level diagnosis. The results demonstrate the viability of our method's performance, together with its effectiveness and robustness, offering a fresh concept for the development of computer-assisted reading systems in clinical cervical screening in the future.

Yuliana et.al [23] developed and preliminarily validated a model based on the Unet network and SVM

to classify cervical lesions on colposcopy images. The Intel & Mobile ODT CC Screening public dataset and a private dataset from an Ecuadorian public hospital were both used as image sets. The Unet and SVM models, respectively, segregated and categorise the cervix lesions or regions of interest. The capability of the CAD system to forecast the risk of CC was assessed. a 65% precision and an 80% accuracy. Sensitivity, specificity, and accuracy of the classification results were 70%, 48.8%, and 58%, respectively. The CAD system needs to be improved, but in a setting where women have limited access to doctors for the diagnosis, follow-up, and treatment of CC, it might be acceptable. Better performance is feasible by exploring alternative DL techniques with large datasets.

Madhura Kalbhor et.al [24] presents the methodology for CC prediction based on pap smear images. Deep neural network models that have already been trained are utilised to extract features, and various ML models are taught on those features. Four pre-trained models, including Alexnet, Resnet-18, Resnet-50, and Googlenet, are fine-tuned for feature extraction in the proposed methodology before various ML algorithms are used. With the Alexnet pre-trained model, the algorithms have outperformed this straightforward logistic regression with the greatest accuracy of 95.14%.

For the detection of CCs various DL methods are been utilized. The methods in use may be complicated and the performance was around 90%. Though such methods perform better, there is a need of improvement in efficiency. To improve the performance and to make a model more robust, an approach needs to be recommended so that the CC can be predicted properly which results in lower mortality rates as it is the second most cause of death due to cancer. To overcome the issues of the existing techniques, a method called CNN-trans-attention layer net is recommended which is a DL approach used for the accurate diagnosis of the CCs in women.

3. PROPOSED SYSTEM

To overcome the drawbacks in the methods in use and to diagnose cervical cancer precisely, a deep learning approach called cervical cancer detection via CNN-trans-attention layer net is recommended for the accurate diagnosis of the disease. Initially the input data is pre-processed by which the noises are removed and the model's robustness is improved followed by feature extraction by which the shape, texture, and colour-based features are extracted from the pre-processed images and the feature selection is done using Fisher's score and mutual information. Finally, classification is done by combining Convolutional Neural Networks (CNNs) with Transformer-based self-attention mechanisms, enabling it to effectively capture intricate patterns in cervical images. Figure 1 shows the block structure of the recommended methodology.

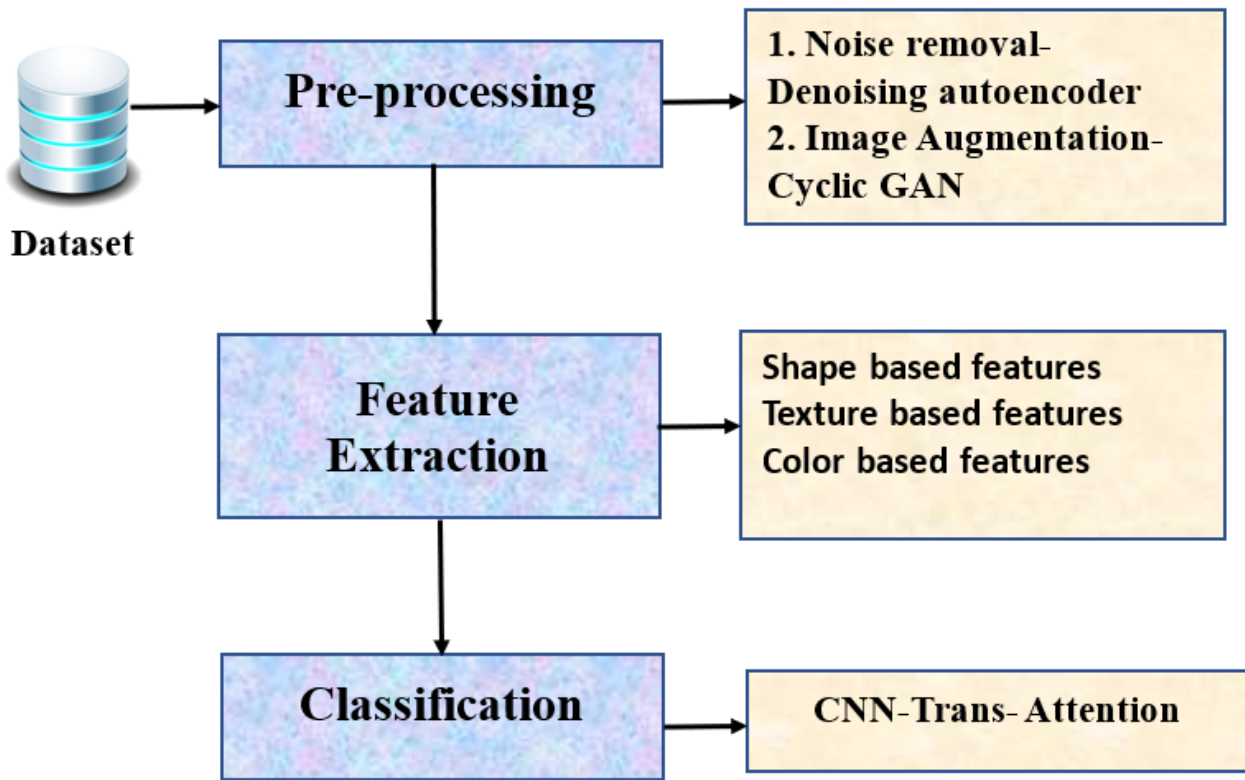


Fig 1: Structure of the recommended methodology.

3.1. Pre-processing

The initial step is the Pre-processing. The data pre-processing is a significant step for transforming the raw dataset into a most suitable format. It involves noise removing enhances the quality of the images. Better preprocessing results lead to superior classification accuracy. The dataset may contain some redundant information and noise signal. Consider a dataset as (X, Y) , where x is data, y is corresponding class, f is features, N is number of instances. The pre-processing is done by means of Noise removal and Image Augmentation.

3.1.1. Noise Removal

Image denoising is the process of eliminating distortion from a noisy image in order to recover the original image without affecting its features and structural integrity. To make the dataset clearer and better suited for further analysis, denoising autoencoders will be used to eliminate noise from the photos. A more reliable version of the common autoencoders is denoising autoencoders. They are trained using samples that have some noise added, but they have the same structure as a standard autoencoder. So, we convert these noisy samples into their pure form. An Auto Encoder is a neural network, which is trained to output the same vector as its clean input vector. The network learns parameters to minimize the squared error between its outputs and target input features. Denoising

autoencoder (DAE) is an extended version of an AE, where the input vector is corrupted by noise, but DAE is still trained to output a clean vector. DAE is a neural network that learns the mapping between noisy and clean output. Figure 2 shows the structure of denoising encoder.

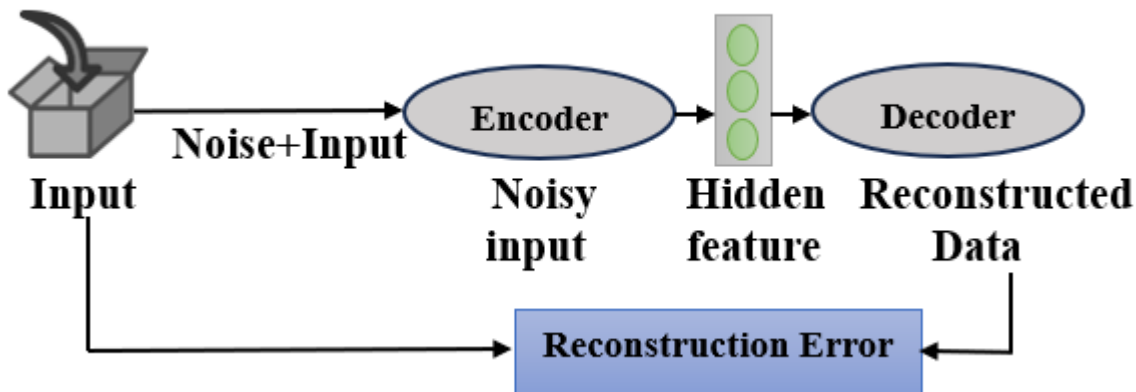


Fig 2: Denoising autoencoder

Let x be the input and x' as a corrupted data. The DAE perturbs the input to x' and maps it to the hidden representation through encoding is given as,

(1)

Where w represents the wights and bias, σ is a nonlinear function (sigmoid)

Then, we reconstruct from h using the decoding function. The output is a linear function and it is computed as,

(2)

The denoising autoencoder is trained by minimizing a reconstruction loss, typically the mean squared error (MSE), between the clean data and the reconstructed data. The loss function is defined as:

(3)

3.1.2. Image Augmentation

In order to give large-capacity learners more illustrative training samples, the size of the training set has been increased using data augmentation approaches. Cyclic GAN will be used in this model to enhance the dataset. By adding changes to the photos, this augmentation process will expand the dataset and strengthen the model's robustness. Generator and discriminator are the two main components of generative adversarial models (GANs). Two GANs make up a cyclic GAN, giving it a total of two generators and two discriminators. Discriminators will distinguish between the generated images and the real images after generators learn the mappings of two domains. By learning the

feature distribution of two domain images, such as colour style translation, Cycle GAN is able to perform image-to-image translation. The loss function of cycle GAN is,

(4)

Where λ represents the weight coefficient. Figure 3 shows the structure of Cyclic GAN

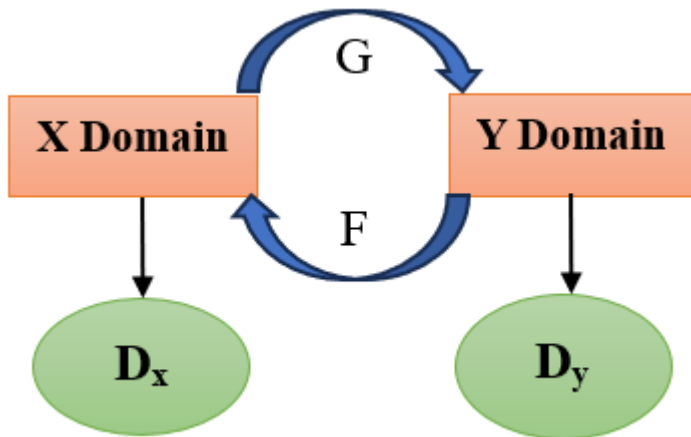


Figure 3: Cyclic GAN

1. : input the original domain slice into the generator to generate slice with domain colour features, and the discriminator judges whether slice belongs to the domain. The loss of the domain GAN is

(5)

G represents the generator from the domain to the domain, represents the discriminator, generated false sample in Y domain. The goal of generator G is to minimize the and the objective of discriminator is to maximize it, so

(6)

2. : input the original domain slice into the generator to generate slice with domain colour features, and the discriminator judges whether slice belongs to the domain. The loss of the domain GAN is

(7)

F represents the generator from the domain to the domain, represents the discriminator, generated false sample in X domain. The goal of generator F is to minimize the and the objective of discriminator is to maximize it, so

(8)

3. is a key component of the loss function. It ensures that the translated images maintain a certain level of consistency when you map them back to the original domain. The loss helps enforce that the

cycle of translations remains close to the original input data. It is given as,

(9)

The original slice of domain and the restored slice of the domain must be same, but there will be the difference between and given as and similarly the difference between and from the domain is given as .

(10)

Where represents norm which measures the absolute differences between corresponding pixels in the two images, represents the expectation over the distribution of real images from Domain X. By minimizing this cyclic consistency loss during training, the model learns to produce translations that are both realistic and maintain the essential characteristics of the original images.

3.2. Feature Extraction

The next step in the process is the feature extraction. Feature extraction is the process of turning a set of features from the pre-processed data. It involves extracting shape, texture, and colour-based features. The detailed explanation is as follows.

3.2.1. Shape Features:

Shape is also considered as an important low-level feature as it is helpful in identification of real-world shapes and objects. Methods such as contour analysis, edge detection, and morphological operations will be used to extract shape information from the images.

3.2.1.1. Contour analysis

The process of obtaining data from the edges of objects or regions inside an image is known as contour analysis in shape features. In order to extract and characterise objects or regions of interest, structural approaches in image processing analyse the forms and structures inside a picture. Graph the contour as $G(V, E)$, where V denotes the components of the image (region, pixel), and E denotes the collection of edges. To locate subparts or sections inside the shape, use graph clustering algorithms like spectral clustering or modularity-based clustering. Clusters stand in for the shape's many structural elements.

3.2.1.1.1. Spectral clustering

It is a powerful technique that uses the eigenvalues and eigenvectors of a graph's Laplacian matrix to perform clustering. Calculate the Laplacian matrix L of the graph,

The Unnormalized Laplacian is given as,

(11)

where D is the degree matrix (diagonal) and W is the weighted adjacency matrix.

The Normalized Laplacian is given as,

(12)

Where I indicate the identity matrix

The eigenvalues of the Laplacian matrix L are real numbers. They are typically arranged in ascending order: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_N$. The number of zero eigenvalues ($\lambda_1 = 0$) corresponds to the number of connected components in the graph. The non-zero eigenvalues provide valuable information about the graph's structure, connectivity, and spectral properties.

3.2.1.1.2. Modularity based clustering

It identifies the communities within a network by optimizing a measure called modularity. For a given partition of nodes into communities, the modularity Q is calculated using the formula:

(13)

Where w_{ij} is the edge weight between nodes i and j , k_i and k_j are the degrees of nodes i and j , m is the total edge weight, C_i and C_j are the communities to which nodes i and j belong, and δ_{ij} is the Kronecker delta, which equals 1 when $i = j$ and 0 otherwise.

3.2.1.2. Edge detection

It refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. For edge evaluation, it generates features that preserve information on gradient orientation and amplitude. This method counts instances of gradient orientation within a specific area of an image.

(14)

(15)

Where r is rows, c is columns, I is image

Once G_x and G_y are calculated, G of a pixel is using the equations shown below to determine the value.

Magnitude (16)

Orientation, (17)

3.2.1.3. Morphological operations

Morphology is a broad range of image processing techniques that manipulate images according to their shapes. A structuring element is added by morphological processes to an input image to produce an output image of the same size. When performing a morphological operation, each output pixel's value is determined by comparing it to its nearby neighbours in the input image. Every pixel in the original image has the structural element moved across it to create a pixel in the newly processed image. The morphological process used determines the value of this new pixel. Erosion and dilation are the two most often employed operations.

3.2.1.3.1. Dilation

It expands the boundaries of objects in a binary image. It is used to fill gaps, connect broken structure and make the objects more robust for processing. Dilation of an image C by a structuring element D is defined as

(18)

where (x, y) are the coordinates of the image pixels, and (i, j) are the coordinates within the structuring element D .

3.2.1.3.1. Erosion

Erosion shrinks the boundaries of objects in a binary image. It is used to remove noise, separate overlapping objects, and extract the core structure of objects. Erosion of an image A by a structuring element B is defined as

(19)

3.2.2. Texture based features

Texture is the most important feature for many types of images that appear everywhere in nature such as medical images and sensor images and so on. Texture analysis techniques called Gray-Level Co-occurrence Matrix (GLCM) will be applied to capture textural details present in the images. Consider an image I of size $K \times K$, elements of a $G \times G$ grey level co-occurrence matrix for a displacement vector is given as,

(20)

There are various texture features specific by GLCM Entropy, Contrast, Correlation, Energy and Homogeneity are explained below.

3.2.2.1. Entropy

A statistical measure of randomness be utilized to distinguish the texture of an input image. It is given as,

(21)

3.2.2.2. Contrast

It calculates the density contrast between pixels and adjacent pixels to the whole image and it is given as,

(22)

3.2.2.3. Correlation

The function of this scale is to measure the probability of the specified of the specified pixel pairs

(23)

3.2.2.4. Energy

It is the summation of squared elements in GLCM. It is also known as the angular second moment. IT is given as,

(24)

3.2.2.5. Homogeneity

It is used to measure the approximation of the distribution of elements in the GLCM to the GLCM diagonal which is defined as,

(25)

indicates the number of gray-level co-occurrence matrices in GLCM, indicates the pixel at location

3.2.3. Colour-based features

The distribution or associations of the colour information in the segmented objects are captured by colour histograms. Due to its intuitive nature relative to other qualities and more significant information, ease of extraction from the image, and the way the histogram distributes collars using a series of boxes, colour is the most prevalent and commonly utilised feature. Even though the phrase "colour histogram" is more frequently associated with three-dimensional colour systems like RGB or HSV, it may be constructed for any type of colour space. The term intensity histogram may be used instead for monochromatic images. The histogram will be used as a model for the probability distribution of the intensity levels in the statistically based histogram features that we will take into consideration. These statistical features provide us with information about the characteristics of the intensity level distribution for the image. We define the first-order histogram probability, as:

(26)

Where represents total pixels in the image, represents total pixels at grey level g

3.2.4. Feature selection using Fisher's score and mutual information

Fisher score is an efficient approach to feature dimension reduction of data [25]. The main purpose is to find the feature subset such that in a dataspace spanned by the selected features, and maximize the distance between the data points in the same class. In particular, given a dataset with respect to c different classes, the Fisher score of the feature is computed as

(27)

is the between-class scatter of the feature, and it is computed as,

(28)

is the within-class scatter matrix of the feature with respect to class, and it is given as,

(29)

Where represents the number of samples in the class, represents the mean of the feature in the class, is the mean of the feature in, and denotes the value of the feature for the sample in the class. These algorithms will collectively determine the most relevant features for the classification task.

The traditional Fisher score model calculates the score of each feature to compute the feature of the multilabel datasets Fisher model based Mutual information, is utilized. Mutual information measures the statistically independent relationship between certain features and classes. Then the mutual information between the feature and label is given as,

(30)

Where is the probability that both , are included in the training set, is the probability that the training set contains , is the probability that a given class in the training set belongs to , is the probability of being included in .

If and , then for any feature and any two labels , the score of each feature is defined as,

(31)

Where denotes the balance coefficient, denotes the number of samples in class, represents the value of feature on sample , indicates the average value of in class and indicates the number of classes.

(32)

represents that when not hit, for hitting labels the frequency parameter is given as,

(33)

Where indicates the number of hitting labels, indicates the value of label to sample , denotes the number of samples, denotes the number of labels. This will collectively determine the most relevant features for the classification task.

3.3. Classification

The cervical cancer classification is presented in the following section. The classification is used for differentiating abnormalities based on features selected in the previous stage. For cervical cancer detection, a novel architecture combining Convolutional Neural Networks (CNNs), transformers and attention mechanisms will be implemented at the layer level. This architecture will leverage the selected features to classify images effectively. Figure 4 shows the block illustration of CNN-Trans-Attention model.

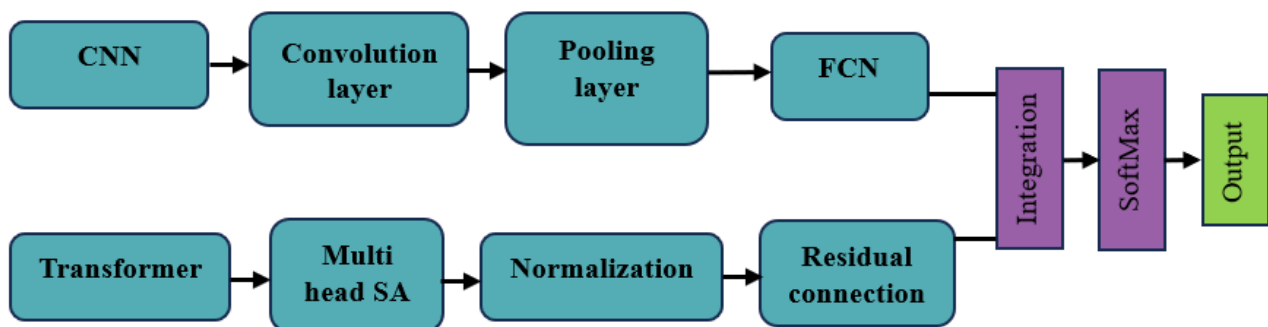


Figure 4: Block illustration of CNN-Trans-Attention model

3.3.1. CNN

It is successful in many image processing applications including medical image analysis. So, CNN-based systems were proposed to detect cervical cancer.

3.3.1.1. Convolution layer (CL):

It is significant part in CNN. In every CL, the i/p cube is convoluted with multiple learnable filters, resulting in generating feature maps. If X be the input cube and its size are $m \times n \times d$, where $m \times n$ refers to the spatial size of X , d is the number of channels, and considering k filters at this CL and the j th filter can be characterized by the weight w_j and bias b_j . The j th o/p of the CL is,

$$, j=1, 2, k \quad (34)$$

$f(\cdot)$ is activation function that is utilized to improve nonlinearity. ReLu is used and it is given as,

(35)

3.3.1.2. Pooling layer (PL):

In case of redundant information in images, PL are periodically added after CL. In this, the spatial size of the feature maps is reduced, and also parameters and computation are reduced. For a $p \times p$ window neighbour denoted as S , the PL o/p is given as,

(36)

where F is total elements in S

a_{ij} is activation value corresponding to the position (i, j) .

3.3.1.3. Fully connected layers (FCL)

After PL, the feature maps are flattened and fed to fully connected layers. This layer changes the previous layers 2D structure features into a predefined one-dimensional feature vector. In a traditional neural network, the FCL are used for reorganizing feature maps, more profound and abstract features can be extracted. into an n -dimension vector.

(37)

where x is input, y is Output, W is Weight and b is bias

3.3.2. Transformer

Transformers are a type of deep learning model used for various tasks that process the entire input data at once, capturing context and relevance. They utilize a mechanism called “self-attention” to process sequential input data.

3.3.2.1. Multi Head self-attention mechanism

The transformer architecture’s fundamental element is the self-attention mechanism. It calculates a weighted sum of the input data, with the weights determined by how similar the features of the input are to one another. This enables the model to give the pertinent input features more weight, which aids in capturing more accurate representations of the input data. Self-attention, then, is a computational primitive that quantifies paired entity interactions and enables a network to understand the hierarchies and alignments contained in incoming data. For visual networks to acquire greater resilience, attention has been shown to be a crucial component. Consider a sequence of entities $x_1, x_2, \dots, x_n$, where d is the embedding dimension to represent each entity. The goal of self-attention is to capture the interaction amongst all entities by encoding each entity in terms of the global contextual information. This is done by defining three learnable weight matrices to transform Queries Q , Keys K , Values V where $Q = xW_Q$, $K = xW_K$, and $V = xW_V$. The input sequence is projected onto these weight matrices to get Q , K , and V . The output of

the self-attention layer is,

(38)

For a given entity in the sequence, the self-attention basically computes the dot-product of the query with all keys, which is then normalized using SoftMax operator to get the attention scores. Each entity then becomes the weighted sum of all entities in the sequence, where weights are given by the attention scores

In order to encapsulate multiple complex relationships amongst different elements in the sequence, the multi-head attention comprises multiple self-attention blocks ($h=8$). Each block has its own set of learnable weight matrices ; where $=0 \dots$. For input , the output of the self-attention blocks in multi head attention is then concatenated into a single matrix and projected onto a weight matrix. The main difference of self-attention with convolution operation is that the filters are dynamically calculated instead of static filters (that stay the same for any input) as in the case of convolution. Further, self-attention is invariant to permutations and changes in the number of input points.

3.3.2.2. Feedforward neural network (FFNN)

Before transmitting the data from the self-attention mechanism, FFNN transformers frequently apply layer normalization and residual connections. Layer normalization aids in training stabilization, and residual connections allow for gradient flow, simplifying DL. Normalization is used for stabilizing and speed up the training of deep neural network. Layer normalization is a specific form of normalization applied at the level of individual layers. For each layer, it normalizes the inputs across the features (along each feature dimension) and independently for each data sample in the mini-batch.

(39)

Additionally, as the network depth increases, it becomes harder for the network to learn identity mappings (i.e., mappings where the output is the same as the input). Residual connections, also known as skip connections or shortcut connections, address these challenges by allowing information from earlier layers to bypass some of the later layers. In essence, they learn residual functions (the difference between the output and input) rather than trying to directly learn the desired output. Given an input X and a neural network layer $F(X)$ (representing the transformation learned by the layer), the output of a residual block with a skip connection is calculated as:

(40)

This combination helps in achieving both stable training and the ability to train very deep networks

3.3.2.3. Integration Layer:

This layer integrates the outputs from the Convolutional Layers and Transformer Encoder Blocks.

3.3.2.4. Fully Connected Layers:

After integration, the combined features are often passed through one or more fully connected layers.

3.3.2.4. Output Layer

Class probabilities are generated by the final output layer using an activation function, commonly SoftMax. The number of classes in your classification issue corresponds to the number of neurons in this layer. It creates a probability distribution over the classes using a vector of raw scores or logits as input. The chance that a given data point belongs to each class is calculated by the SoftMax function. The final class label is anticipated to be the one with the highest likelihood.

(41)

Fine-tune the selected model on your cervical cancer dataset. Fine-tuning involves training the model on your specific dataset to adapt it to the classification task.

4. RESULTS AND DISCUSSIONS

Using the dataset taken, the results are evaluated using the performance metrics——. The results are computed in comparisons of the suggested and the existing methods———— by implementing in a —— platform.

4.1. Dataset Description:

—————

4.2. Evaluation metrics:

The performance of the suggested and the existing approaches are done using the performance metrics called Accuracy, Sensitivity, Specificity, Precision, F measure, FNR, FPR and MCC.

4.2.1. Accuracy:

It is the proportion of true forecasts to all i/p Observations. It is calculated using the following formula,

(48)

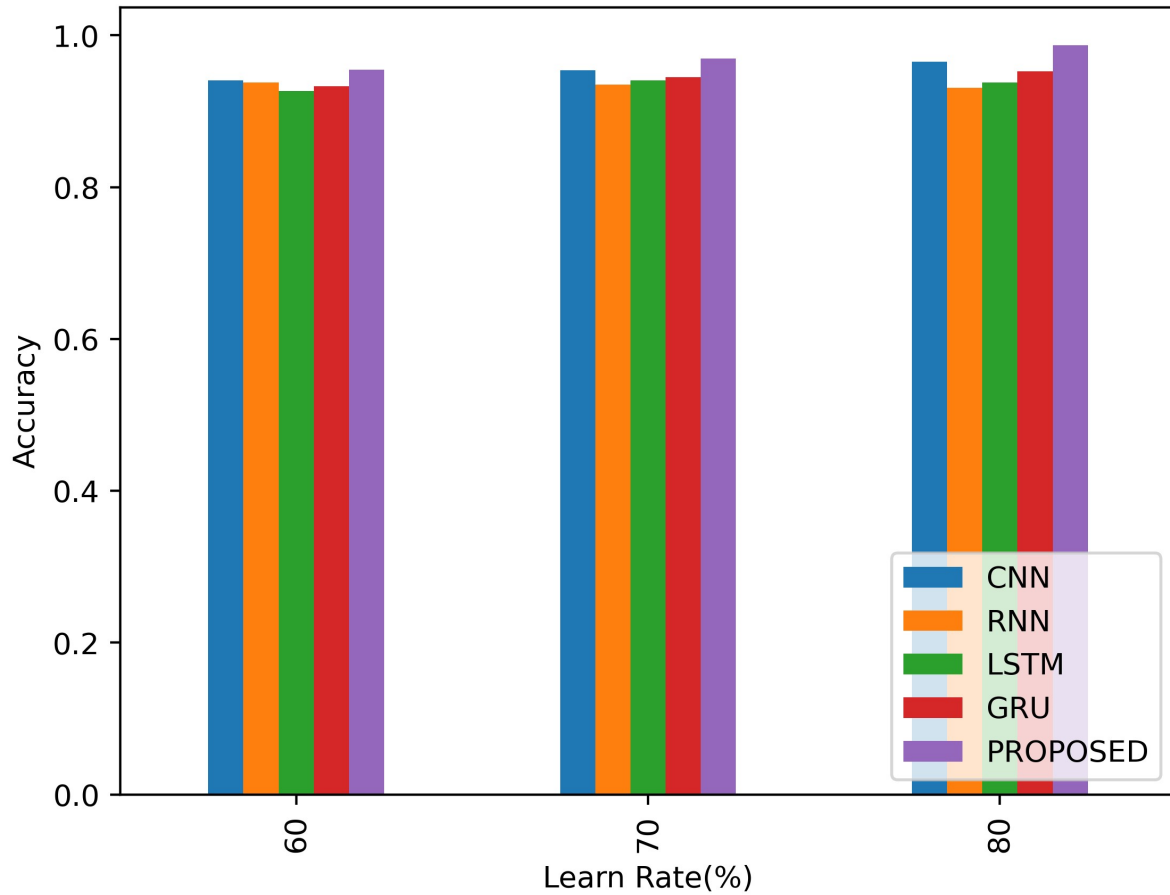


Fig 5: Examination of suggested and existence approaches in terms of Accuracy

Figure 5 shows the examination of the recommended and methods in use in regard to Accuracy. From the graph, the proposed approach has an accuracy of about 98.7% whereas the existing methods CNN, RNN, LSTM and GRU has the accuracy of 96.5%, 93.1%, 93.7, 95.2% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher Accuracy.

4.2.2. Sensitivity

The fraction of real positives that are correctly identified is measured by sensitivity. It is given as,

(49)

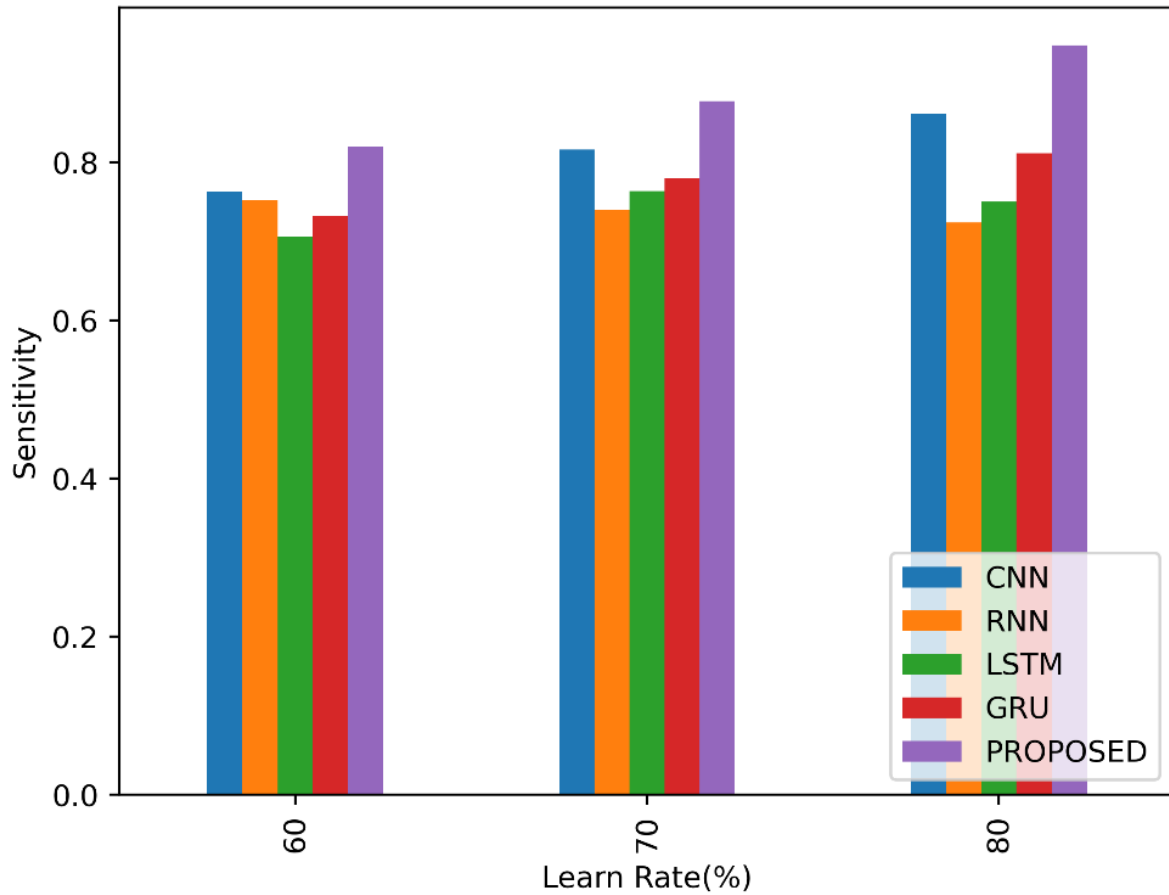


Fig 6: Examination of suggested and existence approaches in terms of Sensitivity

Figure 6 shows the examination of the recommended and methods in use in regard to Sensitivity. From the graph, the proposed approach has a Sensitivity of about 94.8% whereas the existing methods CNN, RNN, LSTM and GRU has the Sensitivity of 86%, 72.4%, 75%, 81.1% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher Sensitivity.

4.2.3. Specificity

The percentage of real negatives that are accurately identified is measured by specificity. It is calculated using

(50)

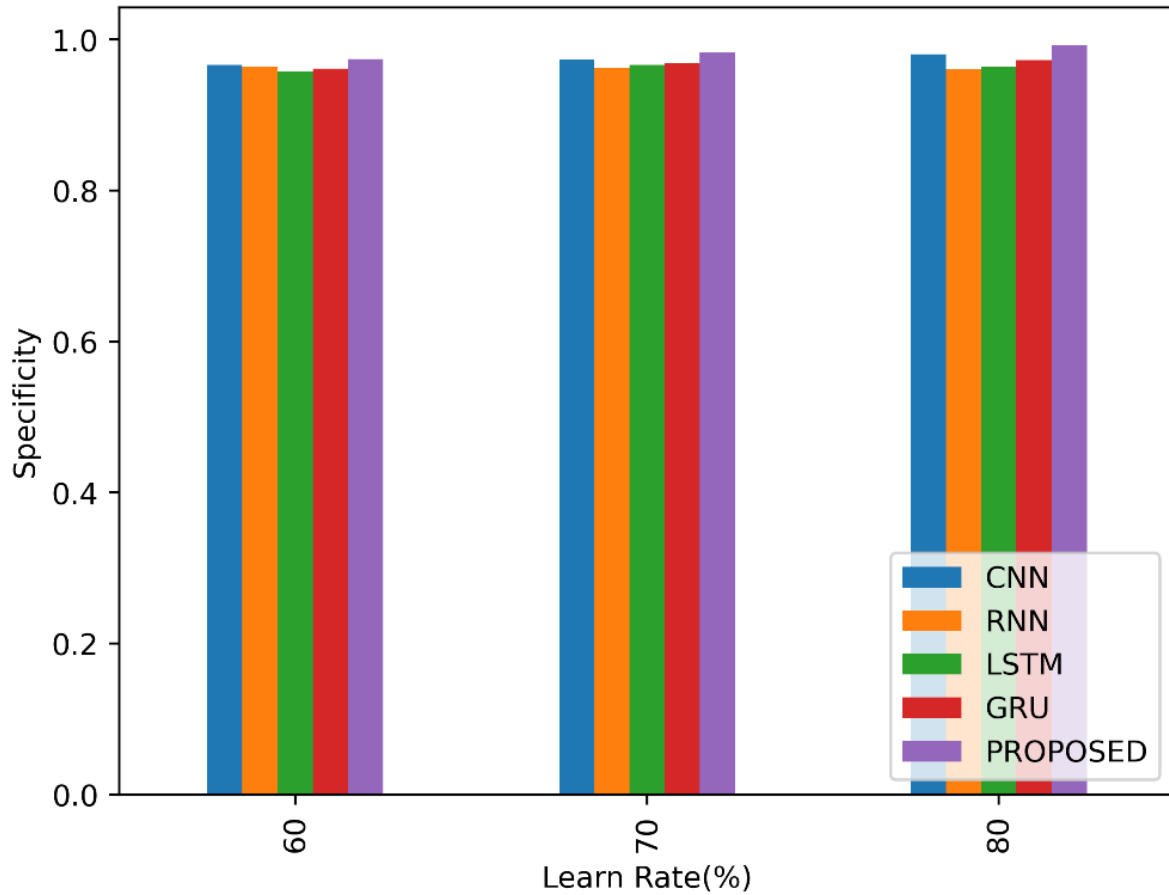


Fig 7: Examination of suggested and existence approaches in terms of Specificity

Figure 7 shows the examination of the recommended and methods in use in regard to Specificity. From the graph, the proposed approach has a Specificity of about 99.2% whereas the existing methods CNN, RNN, LSTM and GRU has the Specificity of 98%, 96%, 96.4%, 97.3% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher Specificity

4.2.4. Precision

How much of a model's positive predictions are actually right is determined by its precision, which is a performance indicator. In order to assess how well what you detect is actually present, precision is important. It is given as,

(51)

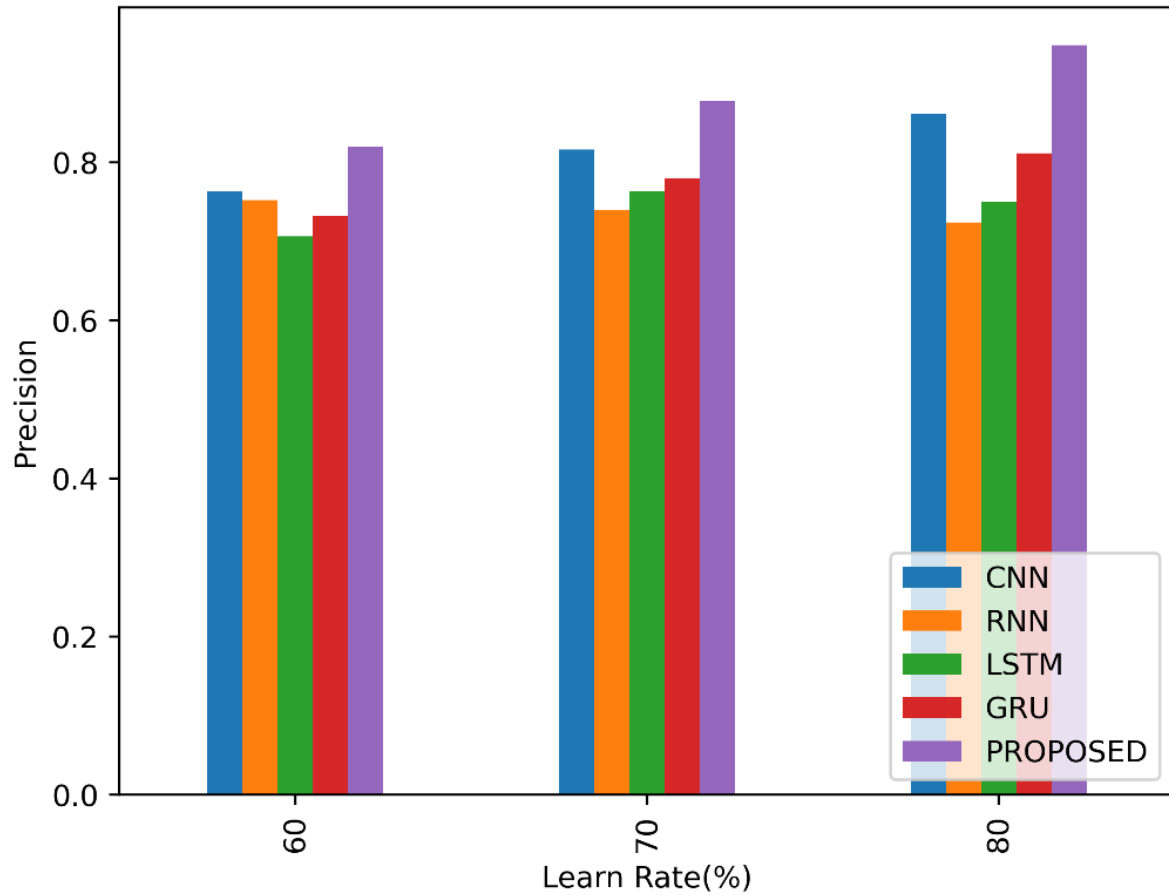


Fig 8: Examination of suggested and existence approaches in terms of Precision

Figure 8 shows the examination of the recommended and methods in use in regard to Precision. From the graph, the proposed approach has a Precision of about 94.8% whereas the existing methods CNN, RNN, LSTM and GRU has the Precision of 86%, 72.4%, 75%, 81.1% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher Specificity.

4.2.5. F measure

A general score for performance evaluation, the F1-score is a combination statistic that combines Precision and recall. It is given as,

(52)

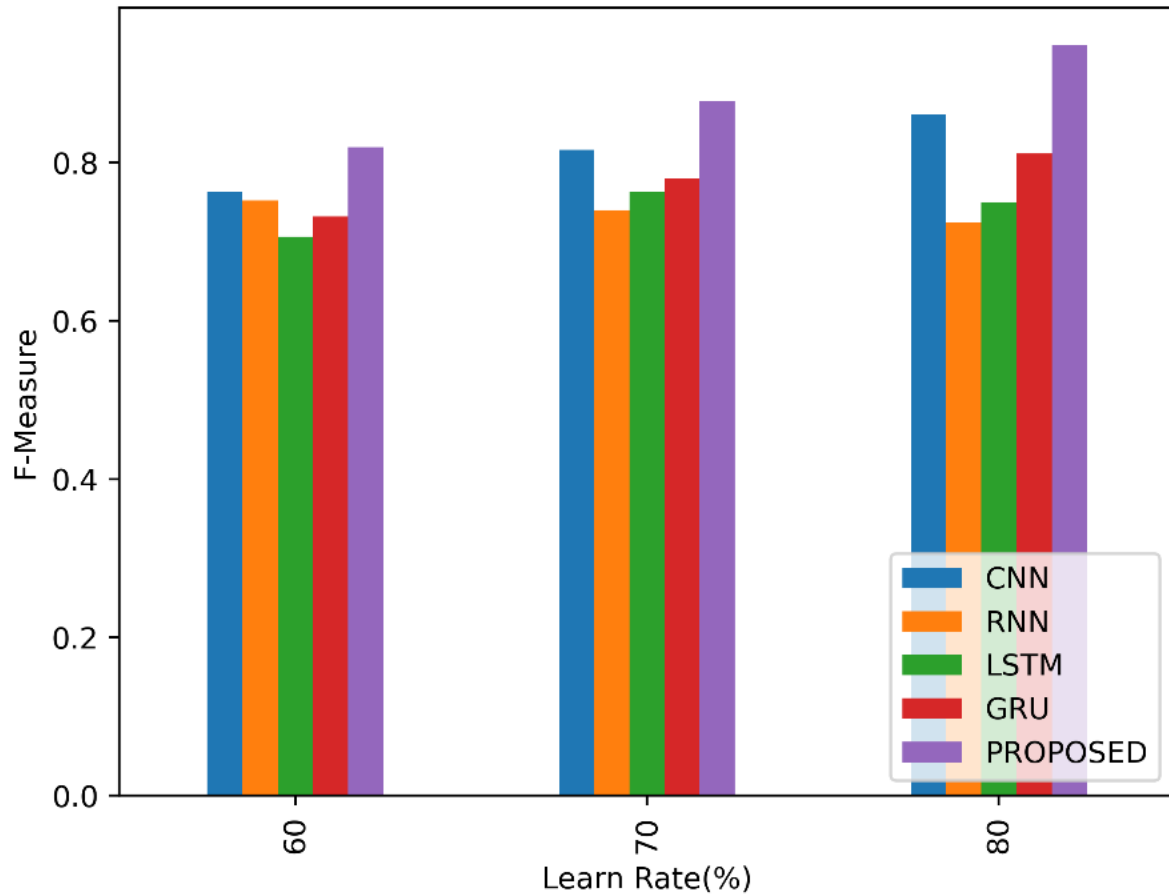


Fig 9: Examination of suggested and existence approaches in terms of F-measure

Figure 9 shows the examination of the recommended and methods in use in regard to F-measure. From the graph, the proposed approach has a F-measure of about 94.8% whereas the existing methods CNN, RNN, LSTM and GRU has the F-measure of 86%, 72.4%, 75%, 81.1% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher F-measure.

4.2.6. False Negative Rate (FNR)

It refers to the values that are actually positive but predicted to negative. It is calculated using the formula,

(52)

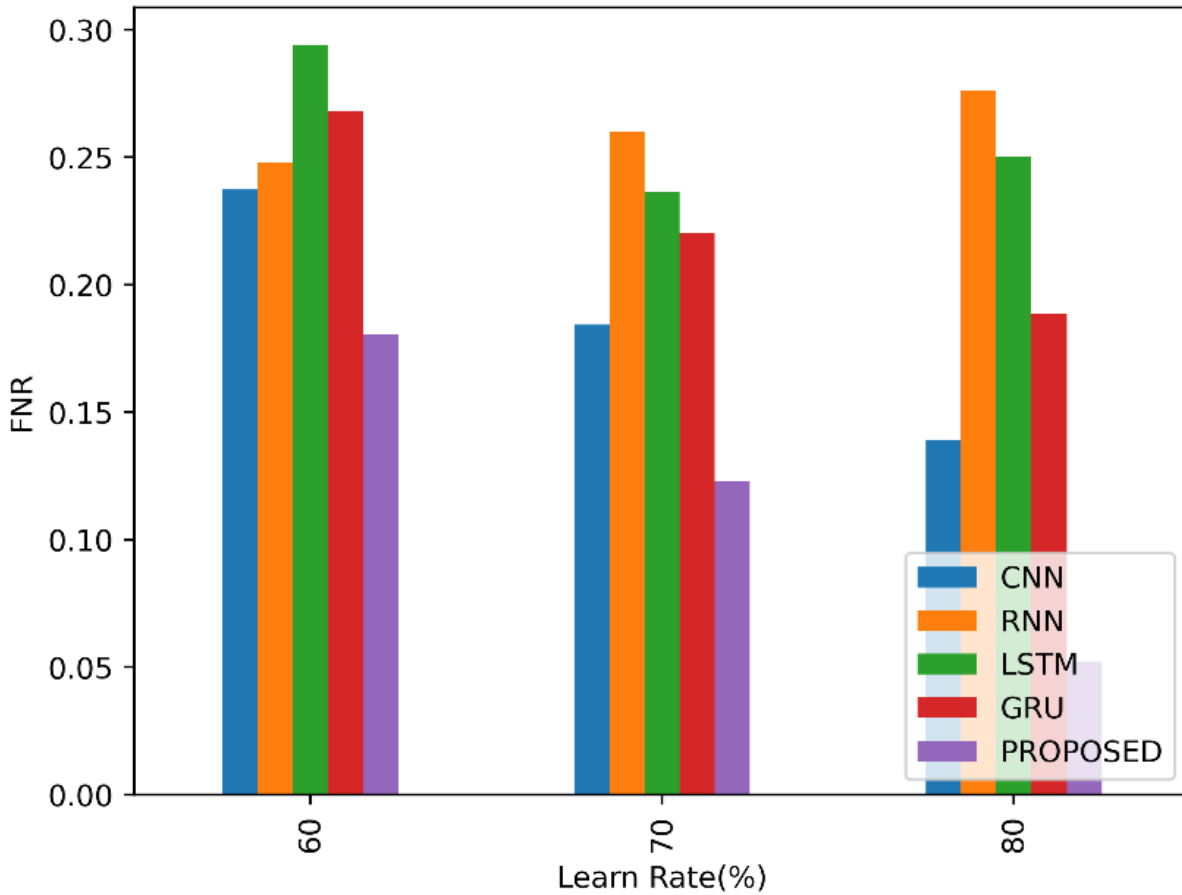


Fig 10: Examination of suggested and existence approaches in terms of FNR

Figure 10 shows the examination of the recommended and methods in use in regard to FNR. From the graph, the proposed approach has FNR of about 5% whereas the existing methods CNN, RNN, LSTM and GRU has the FNR of 13%, 27%, 25%, 18% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a lower FNR.

4.2.7. False Positive Rate (FPR)

It refers to the values that are actually negative but predicted to be positive. It is calculated using the formula,

(53)

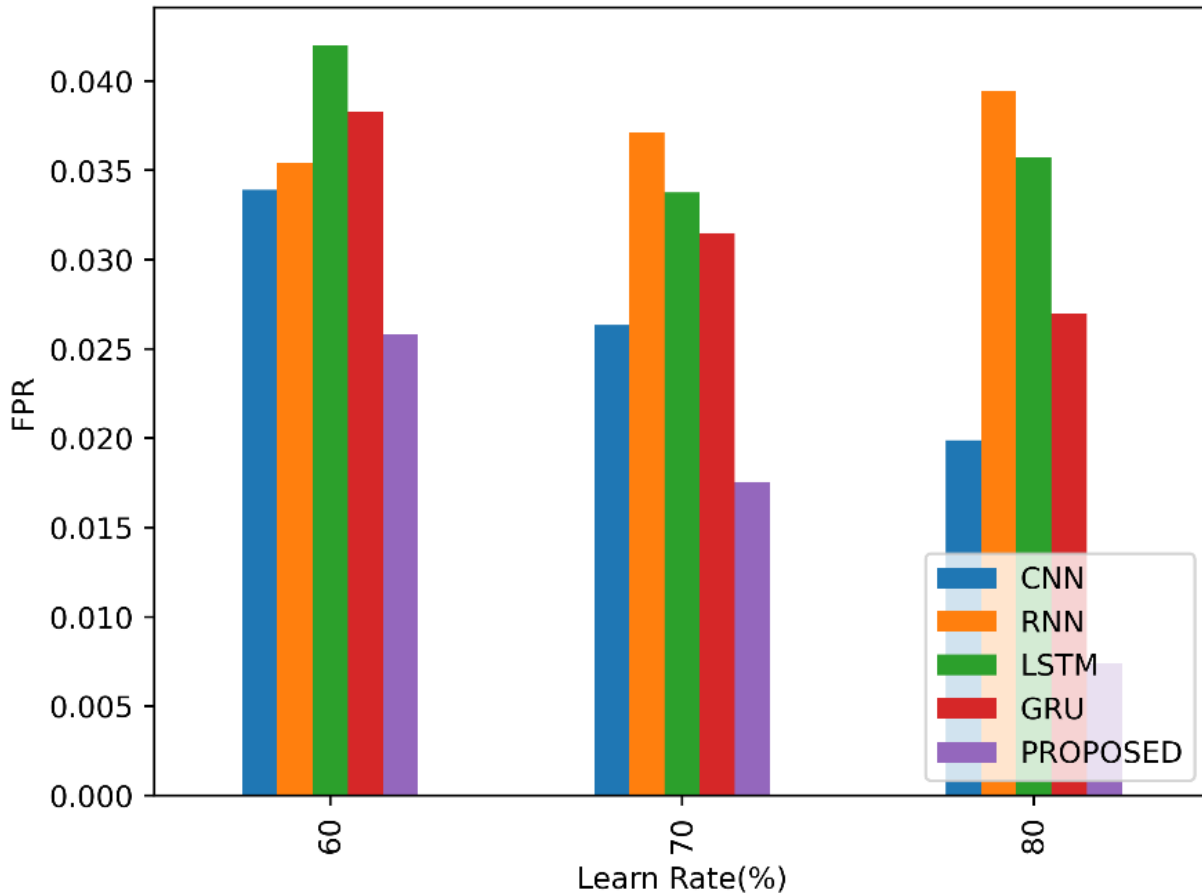


Fig 11: Examination of suggested and existence approaches in terms of FPR

Figure 11 shows the examination of the recommended and methods in use in regard to FPR. From the graph, the proposed approach has FPR of about 0.7% whereas the existing methods CNN, RNN, LSTM and GRU has the FPR of 1.9%, 3.9%, 3.5%, 2.6% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a lower FPR.

4.2.8. Mathew's Correlation Coefficient (MCC)

MCC measures the degree of correlation between expected and actual values. It's stated as,

(54)

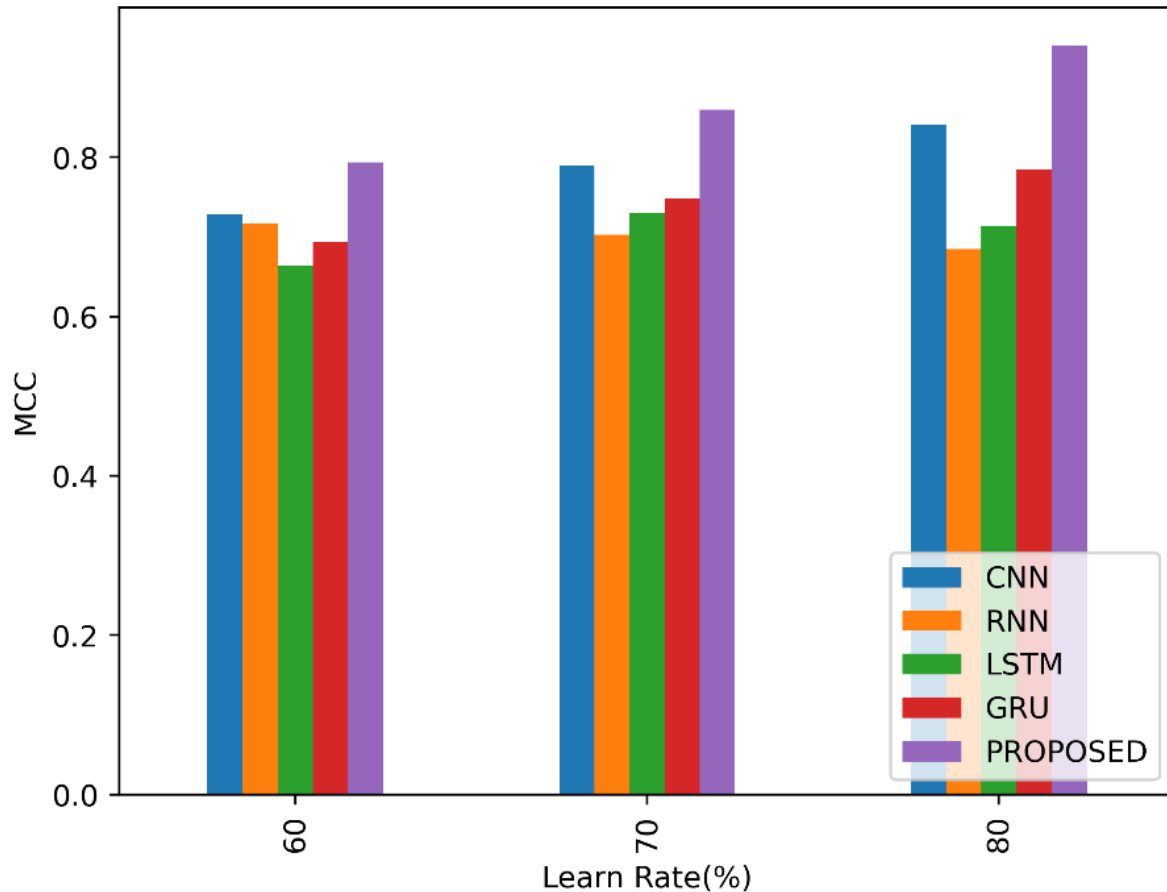


Fig 12: Examination of suggested and existence approaches in terms of MCC

Figure 12 shows the examination of the recommended and methods in use in regard to MCC. From the graph, the proposed approach has MCC of about 94% whereas the existing methods CNN, RNN, LSTM and GRU has MCC of 84%, 68.4%, 71.4%, 78.4% respectively. Thus, from the graphical representation, it is seen that the proposed approach has a higher MCC.

5. CONCLUSION

Early detection of cervical cancer may result in less mortality rates. There exist various methods for the detection of cervical cancer. To overcome the drawbacks in the methods in use and to diagnose this disease precisely, a deep learning approach called cervical cancer detection via CNN-trans-attention layer net is recommended for the accurate diagnosis of the disease. Initially the input data is pre-processed by which the noises are removed and the model's robustness is improved followed by feature extraction by which the shape, texture, and colour-based features are extracted from the pre-processed images. Finally, classification is done by combining Convolutional Neural Networks (CNNs) with Transformer-based self-attention mechanisms, enabling it to effectively capture intricate

patterns in cervical images. The suggested approach has an accuracy of 98.7%, specificity of 99.2%, sensitivity of 94.8%, Precision of 94.8%, F measure of 94.8%. Thus, from the results, it is seen that our suggested approach outperforms well in comparison to the other existing methods.

References

- [1] Thohir, Muhammad, Ahmad Zoebad Foady, Dian Candra Rini Novitasari, Ahmad Zaenal Arifin, Bunga Yuwa Phiadelvira, and Ahmad Hanif Asyhar. "Classification of colposcopy data using GLCM-SVM on cervical cancer." In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 373-378. IEEE, 2020.
- [2] Bouvard, Véronique, Nicolas Wentzensen, Anne Mackie, Johannes Berkhof, Julia Brotherton, Paolo Giorgi-Rossi, Rachel Kupets et al. The IARC perspective on cervical [cancer screening](#)." *New England Journal of Medicine* 385, no. 20 (2021): 1908-1918.
- [3] Kessler, Theresa A. "Cervical cancer: prevention and early detection." In *Seminars in oncology nursing*, vol. 33, no. 2, pp. 172-183. WB Saunders, 2017.
- [4] Jahan, Sohely, MD Saimun Islam, Linta Islam, Tamanna Yesmin Rashme, Ayesha Aziz Prova, Bikash Kumar Paul, MD Manowarul Islam, and Mohammed Khaled Mosharof. "Automated invasive cervical cancer disease detection at early stage through suitable machine learning model." *SN Applied Sciences* 3 (2021): 1-17.
- [5] Cohen, Paul A., Anjua Jhingran, Ana Oaknin, and Lynette Denny. "Cervical cancer." *The Lancet* 393, no. 10167 (2019): 169-182.
- [6] Chakraborty, Sudip, Amar Debbouche, and Valery Antonov. "The role of diagnosis at early stages to control cervical cancer: a mathematical prediction." *The European Physical Journal Plus* 135, no. 10 (2020): 780.
- [7] Shiraz, Aslam, Robin Crawford, Nagayasu Egawa, Heather Griffin, and John Doorbar. "The early detection of cervical cancer. The current and changing landscape of cervical disease detection." *Cytopathology* 31, no. 4 (2020): 258-270.
- [8] V. D. Soni and A. N. Soni, "Cervical cancer diagnosis using convolution neural network with conditional random field," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1749-1754, doi: 10.1109/ICIRCA51532.2021.9544832.
- [9] Khouliqi, Ichrak, and Najlae Idrissi. "Segmentation and classification of cervical cancer." In *2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, pp. 1-7. IEEE, 2020.
- [10] Singh, Sanjay Kumar, and Anjali Goyal. "Performance analysis of machine learning algorithms for

cervical cancer detection." In *Research Anthology on Medical Informatics in Breast and Cervical Cancer*, pp. 347-370. IGI Global, 2023.

[11] Ali, Md Mamun, Kawsar Ahmed, Francis M. Bui, Bikash Kumar Paul, Sobhy M. Ibrahim, Julian MW Quinn, and Mohammad Ali Moni. "Machine learning-based statistical analysis for early-stage detection of cervical cancer." *Computers in biology and medicine* 139 (2021): 104985.

[12] N. Youneszade, M. Marjani and C. P. Pei, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," in *IEEE Access*, vol. 11, pp. 6133-6149, 2023, doi: 10.1109/ACCESS.2023.3235833.

[13] Gupta, Akshat, Alisha Parveen, Abhishek Kumar, and Pankaj Yadav. "Advancement in Deep Learning Methods for Diagnosis and Prognosis of Cervical Cancer." *Current Genomics* 23, no. 4 (2022): 234.

[14] Chitra, B., and S. S. Kumar. "Recent advancement in cervical cancer diagnosis for automated screening: a detailed review." *Journal of Ambient Intelligence and Humanized Computing* (2022): 1-19.

[15] Buskwofie, Ama, Gizelka David-West, and Camille A. Clare. "A review of cervical cancer: incidence and disparities." *Journal of the National Medical Association* 112, no. 2 (2020): 229-232.

[16] Kavitha, R., D. Kiruba Jothi, K. Saravanan, Mahendra Pratap Swain, José Luis Arias Gonzáles, Rakhi Joshi Bhardwaj, and Elijah Adomako. "Ant colony optimization-enabled CNN deep learning technique for accurate detection of cervical cancer." *BioMed Research International* 2023 (2023).

[17] M. Z. H. Ontor, M. M. Ali, S. S. Hossain, M. Nayer, K. Ahmed and F. M. Bui, "YOLO_CC: Deep Learning based Approach for Early-Stage Detection of Cervical Cancer from Cervix Images Using YOLOv5s Model," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-5, doi: 10.1109/ICAECT54875.2022.9807871.

[18] Mehmood, Mavra, Muhammad Rizwan, Michal Gregus ml, and Sidra Abbas. "Machine learning assisted cervical cancer detection." *Frontiers in public health* 9 (2021): 788376.

[19] K. M. A. Adweb, N. Cavus and B. Sekeroglu, "Cervical Cancer Diagnosis Using Very Deep Networks Over Different Activation Functions," in *IEEE Access*, vol. 9, pp. 46612-46625, 2021, doi: 10.1109/ACCESS.2021.3067195.

[20] Elakkiya, R., Kuppa Sai Sri Teja, L. Jegatha Deborah, Carmen Bisogni, and Carlo Medaglia. "Imaging based cervical cancer diagnostics using small object detection-generative adversarial networks." *Multimedia Tools and Applications* (2022): 1-17.

[21] Ming, Yue, Xiyang Dong, Jihuai Zhao, Zefu Chen, Hao Wang, and Nan Wu. "Deep learning-based

multimodal image analysis for cervical cancer detection." *Methods* 205 (2022): 46-52.

[22] Xiang, Yao, Wanxin Sun, Changli Pan, Meng Yan, Zhihua Yin, and Yixiong Liang. "A novel automation-assisted cervical cancer reading method based on convolutional neural network." *Biocybernetics and Biomedical Engineering* 40, no. 2 (2020): 611-623.

[23] Jiménez Gaona, Yuliana, Darwin Castillo Malla, Bernardo Vega Crespo, María José Vicuña, Vivian Alejandra Neira, Santiago Dávila, and Veronique Verhoeven. "Radiomics diagnostic tool based on deep learning for colposcopy image classification." *Diagnostics* 12, no. 7 (2022): 1694.

[24] Kalbhor, Madhura, Swati Shinde, Hrushikesh Joshi, and Pankaj Wajire. "Pap smear-based cervical cancer detection using hybrid deep learning and performance evaluation." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2023): 1-10.

[25] Sun, Lin, Tianxiang Wang, Weiping Ding, Jiucheng Xu, and Yaojin Lin. "Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification." *Information Sciences* 578 (2021): 887-912.

[26] Kang, Zhenping, Yizhe Li, Jie Liu, Cheng Chen, Wei Wu, Chen Chen, Xiaoyi Lv, and Fei Liang. "H-CNN combined with tissue Raman spectroscopy for cervical cancer detection." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 291 (2023): 122339.

[27] AbuKhalil, Tamer, Bassam AY Alqaralleh, and Ahmad H. Al-Omari. "Optimal Deep Learning Based Inception Model for Cervical Cancer Diagnosis." *Comput. Mater. Contin* 72 (2022): 57-71.

[28] Chen, Su. Models of artificial intelligence-assisted diagnosis of lung cancer pathology based on deep learning algorithms." *Journal of Healthcare Engineering* 2022 (2022).

[29] Andjelkovic, Jovan, Branimir Ljubic, Ameen Abdel Hai, Marija Stanojevic, Martin Pavlovski, Wilson Diaz, and Zoran Obradovic. "Sequential machine learning in prediction of common cancers." *Informatics in Medicine Unlocked* 30 (2022): 100928.