A Versatile Detection of Cervical Cancer with i-WFCM and Deep Learning based RBM Classification

[1]Soumya Haridas and [2]Jayamalar T
[1]Department of Computer Science,
[2]Department of Information Technology,
[1,2]Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India.
[1]soumya.smya@gmail.com, [2]jayamalar_it@avinuty.ac.in

Correspondence should be addressed to Soumya Haridas : soumya.smya@gmail.com.

Abstract—One of the most common and curable types of cancer in women is cervical cancer, a common chronic condition. Pap smear images is a common way for screening the cervical cancer. It does not present with symptoms until the disease has advanced stages, cervical cancer cannot be detected in its early stages. Because of this, accurate staging will make it easier to give the patient the right amount of treatment. In this paper proposes the Anisotropic Diffusion Filter has been used to improve the Pap smear image by removing noise and preserving the image's edges. The contrast of a Pap smear image has been enhanced using Histogram Equalization. The enhanced image has been segmented using Improved Weighted Fuzzy C-means clustering to make it easier to identify the effective features. As a result, the effective features are extracted from the segmented region and used by a Restricted Boltzmann Machine classifier based on Deep Learning to classify the cancer. The performance of the proposed cervical cancer detection system can be measured in terms of sensitivity, specificity, F-measure and accuracy. The performance measures for the proposed system can be achieves 95.3% accuracy, 88.6% specificity, 89.13% precision, 88.56% recall, and 89.7% F-measure respectively. Based on simulation results, the proposed method performs better than conventional methods such as RDVLNN, Random Forest (RF), Extreme Learning Machine (ELM), and Support Vector Machine (SVM) for detecting cervical cancer.

Keywords—Anisotropic Diffusion Filter, Histogram Equalization, Improved Weighted Fuzzy C-Means Clustering, RDVLNN, Random Forest (RF), Extreme Learning Machine (ELM), Support Vector Machine (SVM).

I. INTRODUCTION

Cervical cancer is the fourth most prevalent disease in women overall, as reported by the World Health Organization (WHO), with an anticipated 604,000 new cases and 342,000 fatalities in 2020. Almost 90% of all new cases and fatalities worldwide in 2020 took place in low- and middle-income nations [1, 2]. Early identification through routine examinations is crucial because cervical cancer has a long latency period and starts out without any obvious symptoms. Regardless of whether it spreads to other regions of the body, cancer is a condition in which the body's cells develop rapidly. It is typically named for the part of the body where it first appears [3-5]. Cancer that starts in the cervix is referred to as cervical cancer [6, 7].

According to estimates, more than 500,000 women had cervical cancer in the world in 2018, and more than 300,000 of those women passed away from the disease. Nearly all (99%) incidences of cervical cancer are linked to high-risk human papillomavirus (HPV) infection, an incredibly common virus transferred through sexual contact. As a result, screening tests and receiving an anti-HPV vaccine may help to prevent cervical cancer.

Moreover, a Pap smear test is typically used to identify cervical cancer. It is a quick, painless screening test for uterine cervix cancer or pre cancer. Moreover, the frequent Pap test procedure reduces the incidence rate of cervical cancer. An effective segmentation and classification technique is employed to determine the cervical cancer's real stage. Under a microscope, the extracted cervical cells are examined in order to manually classify the aberrant and normal cervical cells. However, this method is vulnerable to high percentages of false positives since cell sorting is sensitive to human error. Pathologists can only count 4 to 5 slides every day using this method, which is quite economical. It is

challenging to carry out the process more quickly because of the cytoplasm and nucleus that are present in the cell structure. Two key procedures are used in every computer-aided cervical cancer screening system: segmentation and classification.

Motivation

The optimization algorithms and machine learning techniques are used in this study to produce superior solutions:

- Establishing a data validation mechanism to enhance cervical cancer prediction performance.
- To effectively locate the multiple cell nucleus rather than the single cell nucleus in an image of cervical cancer.
- How to identify key performance indicators for a classifier that outperforms other classifiers

Objective

Our study has the following objectives:

- Evaluate the performance of the developed system of our classifier.
- Explore and Compare different evaluation features for detecting the cervical cancer.
- Select the key performance indicators on the proposed model of cervical cancer identification and also measure the Accuracy, Specificity, Precision, Recall, and F-measure.

Contribution of the Proposed System

- To achieve precision, an Anisotropic Diffusion Filter with Histogram Equalization has been utilized which has used to remove the unwanted noises and preserve the edges of cells.
- The Improved Weighted FCM has been used to segment the area of interest in order to overcome histology images' composite nature and irregular forms.
- Feature extraction has been used to extract textural based features.
- A Restricted Boltzmann Machine has beenintroduced to classify cervical cancer and compare the accuracy with other existing classifiers.
- Standard benchmark measurements have been used in this research tomonitor the efficiency of an automated system in cervical cancer: Accuracy, Specificity, Precision, Recall, and F-measure.

This paper is organized as follows: section 2: Related Works; section 3: Proposed Methodology; section 4: Results and Discussion, and section 5: Conclusion and Future enhancements.

II.   RELATED WORKS

Cervical cells have been studied extensively in this regard. A two-tier problem involves dividing cells into normal and defective cells, and a seven-tier problem is dividing cells into one of the seven groups. Nine features are used to depict the nucleus region, whereas eleven features are used to represent the cytoplasm.

To categorize cancer using Pap Smear Test images, Geetha and Suganya[8] suggested Elman Neural Network (ENN) working with the Teaching Learning Based Optimization (TLBO) algorithm. An input Pap smear image is first transformed from RGB to grey level. The grey level image is smoothed with the Kuan Filter during preprocessing in order to remove undesired noise created (KF). The detected cells from the Pap smear picture have been segmented using the Active Contour Method (ACM). To increase accuracy, features like GLCM, haralick, solidity, form, and other mathematical features are extracted. The ENN-TLBO classification algorithm was used. TLBO is used to obtain the best weights possible during the training period. Performance evaluation was carried out using experimental results, where ENN-TLBO produced good accuracy of 86.6%, outperforming other popular algorithms like SVM and RBF classifiers.

An Additional classification of cervical cells based on a multi-domain hybrid deep learning architecture was proposed by Chuanwang Zhang et al. [9]. By attempting for the first time to classify cervical cells using the multi-domain hybrid deep learning framework (MDHDN), they address the restrictions. The pretrained VGG-19 (Visual Geometry Group-19), a deep convolutional neural network (CNN) includes a hashing layer after the last fully connected layer, extracts cell deep features from multi-domain (time and frequency). Feature selection, clustering, and dimensionality reduction are used to process manually created features for the source photos. Following the output of the category results by the three sub channels of the proposed framework using the SVM classifier, the correlation analysis produces the final cell diagnosis. Findings indicate that the suggested method performs similarly to state-of-the-art models that employ novel structures and have accuracy, sensitivity, and specificity values in the Herlev dataset of 98.7%, 98.2%, and 98.9%, respectively.

Cervical cell multi-classification technique using global context information and attention mechanism was proposed by Jun Li et al. [10]. To categorize cervical cells, they created a convolutional neural network (L-PCNN) that combines information about the whole environment with an attention mechanism. To extract deep learning features, the cell image is forwarded to the upgraded ResNet-50 backbone network. Each convolution block adds a convolution block learning algorithm to instruct the network to concentrate on the cell area, improving the extraction of deep features. After that, the long short-term memory module (LSTM) and pyramid pooling layer are added at the end of the backbone network to combine picture features in various locations. The integration of low-level and high-level features allows the network as

a whole to learn more about regional detail features and resolves the issue of network gradient vanishing. The SIPaKMeD open data set is used for the experiment. The experimental findings reveal that the suggested l-PCNN has accuracy in cervical cell classification of 98.89%, sensitivity of 99.9%, specificity of 99.8%, and F-measure of 99.89%, which is better than other cervical cell classification models, demonstrating the model's efficacy.

Several supervised machine learning methods were examined by Gaurav Kumavat et al. [11] in order to find cervical cancer in its earliest stages. A dataset on cervical cancer was taken from the UCI library and used to train the machine learning model. Using 36 risk indicators and one outcome variable, this dataset of 858 cervical cancer patients was used to compare the various approaches. In this study, six classification methods were used: a random tree, a logistic tree, an XG-boost tree, a Bayesian network, an SVM, and an artificial neural network. To assess the effectiveness and accuracy of the classifiers, all models were trained both with and without a feature selection technique. There were three feature selection techniques used: relief rank I wrapper method (ii), and LASSO regression (iii). With XG Boost's full feature set, the highest accuracy of 94.94% was noted. Also, it has been noted that the feature selection algorithm sometimes outperforms the dataset. However, the drawbacks of prediction studies and models, such as overfitting, lack of interpretability, and simplified, inadequate information, point to the need for additional work to increase the precision, dependability, and utility of clinical outcome prediction.

A unique method for the automatic identification of cervical cancer was presented by Lavanya and Thirumurugan[12] employing modified fuzzy C-means, textural and geometric feature extraction, Principal Component Analysis (PCA), and classification. Despite the uncertainty, modified fuzzy C-means segment the input image into useful sections with promising outcomes. By retaining only, the uncorrelated features, PCA is used to minimize the dimensionality of the data collection and shorten the algorithm's processing time. By using K Nearest Neighbor (KNN) classification with k-fold cross-validation, the images from pap smears are divided into normal and abnormal cells, and the results are compared with those from Fine Gaussian SVM, Linear Discriminant and Ensemble Bagged trees. By evaluating the minimum accuracy, average accuracy, sensitivity, specificity, maximum accuracy, F1-score, and precision, the effectiveness of the suggested approach is evaluated. With minimum accuracy of 94.15%, maximum accuracy of 96.28%, average accuracy of 94.86%, sensitivity of 97.96%, specificity of 83.65%, F1-score of 96.87%, and precision of 96.31% for threefold cross-validation, the experimental findings of the suggested technique demonstrate excellent results.

Research Gap

To automate the system to fill in a number of research gaps from the literature for the following reasons as

- Increase the sensitivity and specificity of the Pap smear test
- Reduce the workload of medical professionals and lab technicians;
- Make cervical cancer screening programs less expensive;
- Decrease the incidence of cervical cancer and mortality rates.

III. PROPOSED WORK

The research background demonstrates that Machine Learning and Deep Learning are increasingly being used to process medical images, particularly to diagnose Cervical Cancer. In order to achieve this goal with improved performance, the present study proposes an optimized cervical cancer segmentation method that makes use of Weighted FCM. The general method for the proposed strategy has been displayed in **Fig 1.**
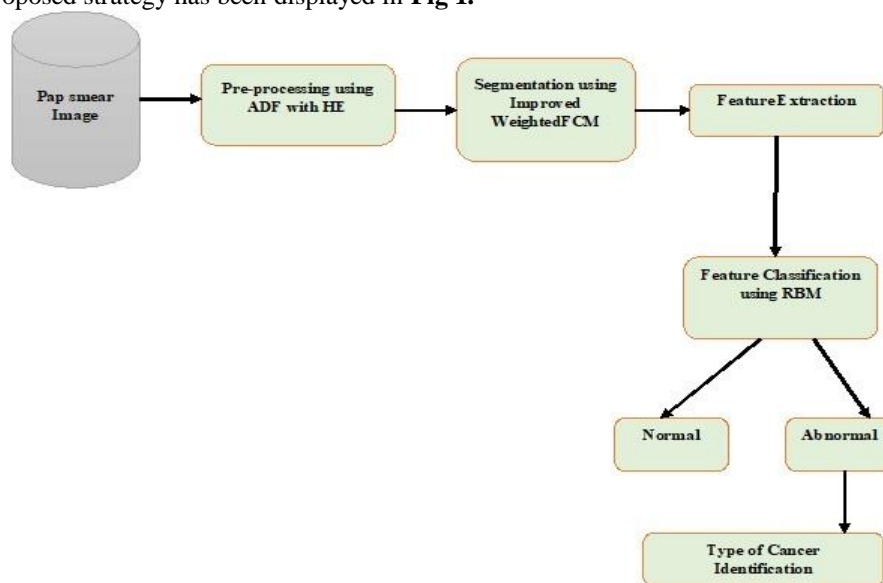

Fig 1. System Architecture of Cervical Cancer Detection

Journal of Machine and Computing 3(3)(2023)

Pre-processing
Operations with images at the lowest level of abstraction are referred to as pre-processing because both the input and the output are intensity images. An intensity image is typically represented by a matrix of image function values (brightness), and these representative images are of the same kind as the original data which is captured by the sensor. The geometric transformations of images, such as rotation, scaling, and translation, are classified as pre-processing methods. The goal of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing.

The process of the Preprocessing in the proposed work is
- Converting the given color image of cervical cancer to Grayscale image transformation.
- Utilize the Anisotropic Diffusion Filter to eliminate any unnecessary noise from the image in order to achieve a high-quality result.

Adaptive Histogram Equalization has been used to make the more enhance image in a very effective manner. The Preprocessing Steps are depicted in **Fig 2.**
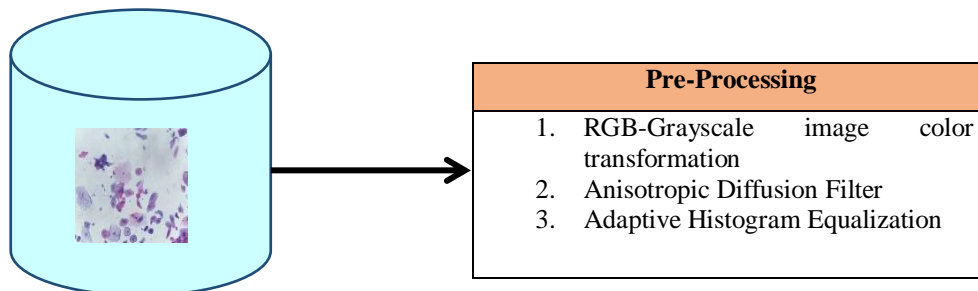

Fig 2. Preprocess

Anisotropic Diffusion Filter
The first and foremost step is to convert the Grayscale image of Pap smear image from the RGB image. In order to enhance the quality of an image, Anisotropic Diffusion Filter has been utilized by removing the noise in the image. Anisotropic diffusion, also known as Perona–Malik diffusion, is a method that aims to reduce image noise without removing significant parts of the image's content [12][13][14], typically edges, lines, or other important details for medical image interpretation. An image generates a parameterized family of successively more blurred images based on a similar process to anisotropic diffusion. Each of the images that come out of this family is a convolution of the image and a 2D Gaussian filter whose width grows as the parameter is changed [15][16].

The Partial Differential Equations (PDE) were used to design the anisotropic diffusion filter, which makes the image diffusion process simpler. This process of diffusion can be extended to include anisotropic diffusion; a collection of bound images is produced by the filter. The superposition of the input image and the filtered content of the input image are used to create the output image.A PDE, which is frequently utilized for noise removal, image edge detection, and detail preservation, is the source of the anisotropic diffusion equation.

The image's edge information can be preserved during denoising by the equation, which can adaptively alter the diffusion coefficient in response to the image's characteristics.
The expression for the diffusion model is

$$\frac{\partial I}{\partial t} = div(c(a,b,t)\nabla I) = \nabla c . \nabla I + c(a,b,t)\Delta I \qquad (1)$$

Where,
$\{\displaystyle \Delta \}\Delta$ denotes the Laplacian, $\{\displaystyle \nabla \}\nabla$ denotes the gradient, $div(...)$ is the divergence operator and $c(a,b,t)\Delta\{\displaystyle c(x,y,t)\}$ is the diffusion coefficient.

For $\{\displaystyle t>0\} t > 0$ the output image is available as $I(.,t)$ $\{\displaystyle I(\cdot,t)\}$, with larger $\{\displaystyle t\}t$ producing blurrier images.

$c(a,b,t)$ Controls the rate of diffusion and is usually chosen as a function of the image gradient so as to preserve edges in the image

$$c(\|\nabla I\|) = e^{-\left(\|\nabla I\|/_K\right)^2} \qquad (2)$$

Contrast Limited Adaptive Histogram Equalization
CLAHE [17] was a step up from AHE. An image is broken up using the CLAHE algorithm into contextual regions. Each contextual region's histogram is generated, and clipping is carried out at a predetermined value. The histogram bins are redistributed with the clipped amount. The problem of AHE's edge-shadowing and over-enhancement is reduced by this

approach. CLAHE has demonstrated its effectiveness in enhancing medical images with low contrast. By redistributing the gray values used, this method makes the images hidden features more visible.

The intensity level of each pixel is compared to the intensity values of its adjacent pixels to determine its ranking. The pixel is then given a new intensity value that is proportional to its rank in the available range. The local area or contextual region, rather than the entire image, is what this method uses to boost contrast.

Segmentation Weighted FCM
Image segmentation is the most common method and analysis in Digital Image Processing, to divide an image into multiple regions based on the characteristics of the pixels [18]. It is the process of separating the foreground from the background, or clustering pixels into regions based on similarities in color or shape.Image segmentation can be used to filter noisy images, find objects in satellite images, perform object detection and recognition tasks, automate traffic control systems, and monitor video.[19][20]

Improved Weighted Fuzzy C-Means (IWFCM) algorithm
The fuzzy c-mean calculation is one of the normal calculations that used to divide the image into different group clusters based on image pixels values. Fuzzy clustering is the most appropriate type of clustering for segmenting medical images. The k-means algorithm can be thought of as the fuzzified version of the Fuzzy C-Means (FCM) algorithm. It is a clustering algorithm that allows data items to be classified according to their degree of membership in each cluster [21][22]. Even though this algorithm reduces the noise with less robust, it make the noisy data into separate cluster. This leads to the poor performance of the FCM.

In order to overcome this poor performance, modified FCM plays uniform contribution of Cluster analysis.

The objective function of FCM is defined as (3)

$$Y_l = \sum_{i=1}^{n} \sum_{j=1}^{c} b_{ij}^l \|x_i - c_j\|^2 \tag{3}$$

where$l$ represents the degree of fuzziness and is a real number greater than 1, bij is the membership degree of the ith datum in the jth cluster, x_i denotes the data points, and c_j is the cluster center. Also, $\|.\|$ represents the Euclidean distance, n is the number of data points, and c denotes the number of clusters. Determination of cluster center is not accurate. So, in order to find the weighting factor and improvement in cluster center, an Improved Weighted Fuzzy C-Means (IWFCM) algorithm is proposed. The proposed IWFCM clustering algorithm still works in the original data space, i.e. prototypes are located in data space, in contrast to the usual method used in FCM. IFCM is particularly well-suited for dealing with incomplete data because it is more resistant to outliers and noise than FCM.

Proposed Improved Weighted Fuzzy C-Means (IWFCM) Algorithm
The proposed IWFCM algorithm is to perform FCM in a higher-dimensional feature space after mapping the input data into it using a nonlinear transform. The objective function that follows is minimized by the Kernel Weighted Fuzzy C-Means. [23]

$$Y\omega = \sum_{i=1}^{k} \sum_{j=1}^{n} b_{ij}^l \|\omega(x_j) - \omega(v_i)\|^2 \tag{4}$$

Where, bij denotes the membership of x_j in cluster i, $\omega(v_i)$ is the center of cluster i in the feature space, and ω is the mapping from the input space X to the feature space F.

Algorithm
KFCM Algorithm
1. Select initial class prototype {Vi} c
2. Update all memberships bijusing the weighted average for cluster center by
3. $u_{ik} = \dfrac{(1 - B(x_k, v_i))^{-\frac{1}{(m-1)}}}{\sum_{j=1}^{c}(1 - B(x_k, x_j))^{-\frac{1}{(m-1)}}}$ $\tag{5}$
4. Obtain the prototype of clusters in the forms of weighted average using the following
5. $v_i = \dfrac{\sum_{k=1}^{n} u_{ik}^m B(x_k, v_i) x_k}{\sum_{k=1}^{n} u_{ik}^m B(x_k, v_i)}$ $\tag{6}$
6. Repeat step 2-3 till stopping criterion is met
7. The termination criterion is$|V_{new} - V_{old}| \le \varepsilon$
8. Where $\|.\|$is the Euclidean norm. V is the vector of cluster centers ε is a small number that can be set by user (here ε=0.01)

Feature Extraction

The process of extracting more specific information from an image is called Feature Extraction. In this stage of the processing, the features extraction process is used to extract the most important features of the segmented cervical cancer area in order to make the diagnosis easier and more accurate. To characterize the cervical cancer and feed the classifier, Feature Extraction seeks to extract features from the Pap smear image. The most basic to the most advanced feature extraction algorithms for diagnosing cervical cancer from Pap smear images were presented in the recent research. Some of the features have been extracted for the identification of lesion image of skin is Geometrical features such as Asymmetry, Diameter, Concavity, Area, perimeter, eccentricity and other features such as Shape, Size, Texture identification [24] GLCM and Haralick Features.

Feature Classification using Restricted Boltzmann Machine (RBM)

A Neural Network known as the Restricted Boltzmann Machine (RBM) simulates a non-directed, symmetrical connection without any intra-layer connections between the visible and hidden nodes. When a set of patterns is given to the network as input, a RBM learns a probability distribution. A Deep Belief Network (DBN) is a deep neural network with many hidden unit layers. Each pair of connected layers in a DBN is a RBM. The data's input is set up in the input layer, and the abstract description of this input is characterized in the hidden layer.

A restricted Boltzmann machine has units connected across layers but no communication within layers. Based on the Hinton's contrastive divergence algorithm which is used to learn the weights of connections between visible and hidden nodes because there are no intra-layer connections [25].

In an RBM, an energy function is defined based on the visible-hidden arrangement of Gaussian neurons.

$$E(v, h) = \sum_i \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_i \sum_j \frac{v_i}{\sigma_i^2} h_j w_{ij} - b'h \tag{7}$$

Where w is the visible –hidden weight matrix, and a and b are the bias vectors respectively.

To calculate the Joint Probability in RBM configuration, $p(v, h)$

$$P(v, h) = \frac{\exp^{-E(v,h)}}{K} \tag{8}$$

Where K is the Partition function can be expressed in

$$K = \sum_{v,h} \exp^{-E(v,h)} \tag{9}$$

Each vector in RBM can be assigned in the Probability as

$$P(v) = \frac{1}{K} \sum_h \exp^{-E(v,h)} \tag{10}$$

The conditional probabilities can be written based on the sigmoid function

$$P(h_j = 1 | v) = \sigma(\sum_i \frac{v_i}{\sigma_i^2} w_{ij} + b_j) \tag{11}$$

$$\text{Where } \sigma(x) = \frac{1}{1 + e^{-x}} \tag{12}$$

and

$$p(v_i = v | h) = \in (v | \sum_j h_j w_{ij} + a_i, \sigma_i^2) \tag{13}$$

Measuring Classification

To estimate the effectiveness of the RBM based cervical cancer detection method, appropriate metrics are utilized in this paper. The F-Measure, Recall, Sensitivity, Specificity, Accuracy, and Precision measures are utilized. For all images, let TP denote the value of the True Positive Rate, FP the value of the False Positive Rate, TN the value of the True Negative Rate, and FN the value of the False Negative Rate.

$$\text{Precision} = TP / (TP + FP) \tag{14}$$
$$\text{Recall} = TP / (TP + FN) \tag{15}$$
$$\text{Specificity} = TN / (TN + FP) \tag{16}$$
$$\text{Accuracy} = (TP+TN)/ (TP+FN+TN+FP) \tag{17}$$

$$\text{F-Measure} = 2*(Precision*Recall)/ (Precision+Recall) \tag{18}$$

IV. RESULTS & DISCUSSION

Dataset Description

SIPaKMeD [26] and Herlev datasets [27] were the two datasets used in the proposed method. The Herlev dataset was utilized for single-cell classification, while the SIPaKMeD dataset was utilized for multi-cell classification. There were 917 images in the Herlev dataset. Classes 1 to 3 are cervical cells that are normal, while classes 4 to 7 are cervical cells that are abnormal. There were 966 images in the multi-cells dataset, and 4049 cells were cropped from these images. The normal, benign, and abnormal stages of the cell were separated.

Experimental Setup

The result of the systematic model is validated using Matlab 2022b simulation tool. The processor includes Intel (R) Core (TM) i5-3210M, CPU@2.5GHz, 2.0GB of RAM. A proposed RBM classification scheme has been evaluated and compared to the existing classification schemes like SVM, RF, ELM and RVDLNN in this section. Images from the MRI of Pap smear image database are initially tested before being trained. MATLAB is used to evaluate the results of the simulation. This evaluation takes into consideration an around 250 sample images. This consists of a training set of 150 images and a testing set of 100 images. The system processing times takes 10mins. **Fig 3** depicts the proposed RBM process in the Pap Smear Image Classification.
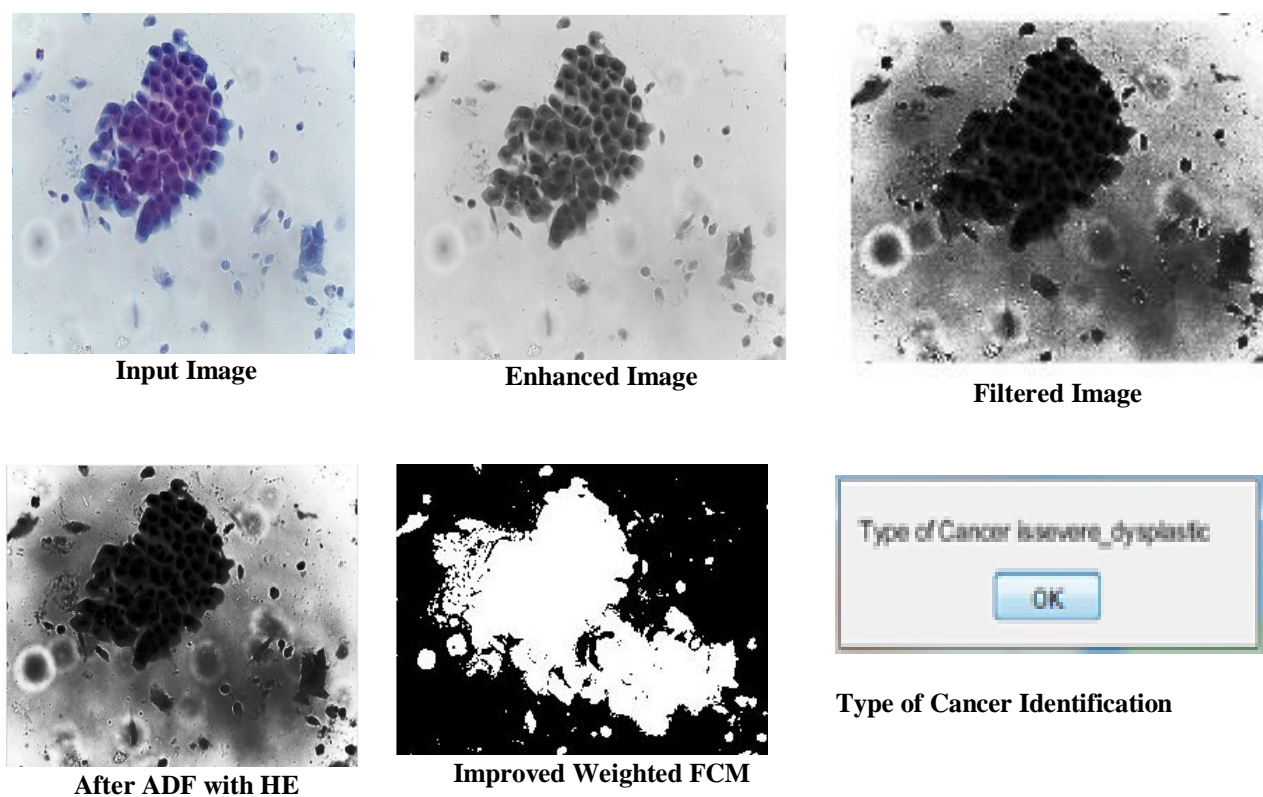

Input Image   **Enhanced Image**   **Filtered Image**

After ADF with HE   **Improved Weighted FCM**   **Type of Cancer Identification**

Fig 3. Proposed RBM Based Results for Input Pap Smear Cancer

Table 1. Training and Testing Accuracy

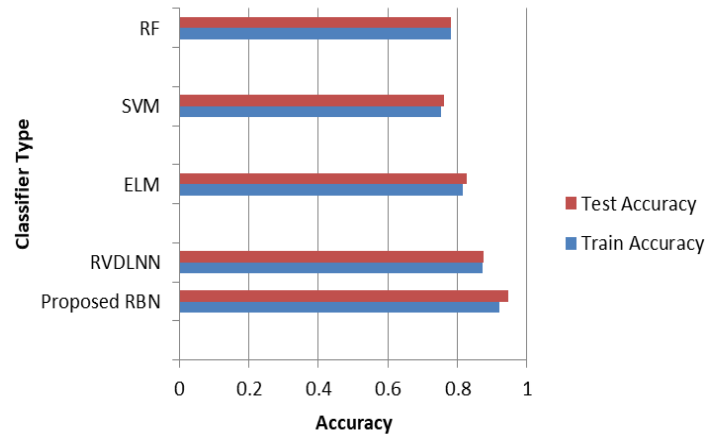| Classifier	Train Loss	Test Loss	Train Accuracy	Test Accuracy
Proposed RBN	0.125	0.126	0.921	0.946
RVDLNN	0.131	0.131	0.872	0.875
ELM	0.143	0.137	0.814	0.827
SVM	0.147	0.21	0.754	0.762
RF	0.162	0.152	0.781	0.782

Fig 4. Training & Testing Accuracy results of Classifiers

Fig 4 depicts the Training and Testing accuracy of all classifiers implemented in cervical cancer Based on the improvement of weight vector calculation in our Proposed RBM, the true positive rate has been improved in **Table 2.**

Table 2. Specificity

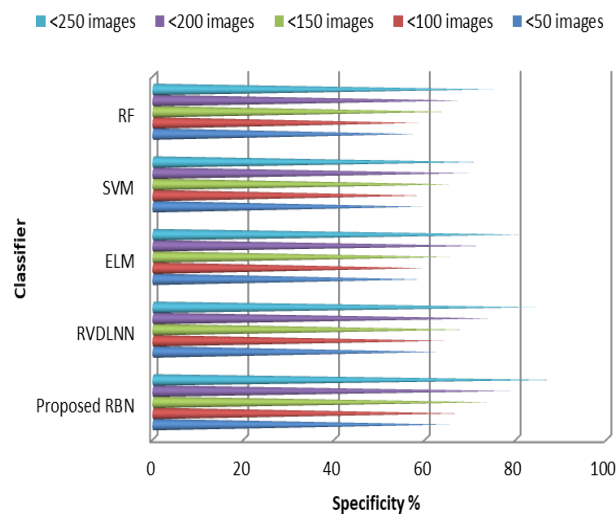| Image Set	Proposed RBN	RVDLNN	ELM	SVM	RF
50	65.12	62.87	59.25	59.25	57.79
100	67.84	64.08	59.75	59.25	58.2
150	74.29	68.96	65.32	65.78	64.82
200	78.54	74.52	72.61	69.12	67.84

Fig 5. Specificity results of Classifiers

The Specificity execution of cervical malignant growth detection plans graphical depiction is showed up in **Fig 5**. It demonstrates that proposed RBM specificity is 4.43%, 7.35%, 16.41% and13.79% higher than that of RVDLNN, ELM, SVM and RF respectively for 250 image dataset. Inferable from true positive rate and true negative results, the specificity of proposed RBN is extended in **Table 3**.

Table 3. Precision

Image Set	Proposed RBN	RVDLNN	ELM	SVM	RF
50	72.57	68.12	61.04	56.12	54.3
100	74.68	67.91	63.81	59.87	57.86
150	74.09	68.25	67.77	64.72	62.06
200	81.63	72.56	74.06	71.66	70.72
250	89.13	78.52	79.15	74.52	72.86

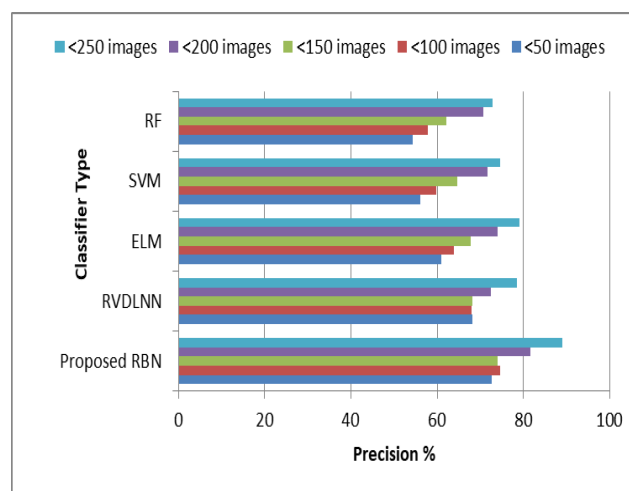**Fig 6.** Precision results of Classifiers

Fig 6 depicts the precise execution of the graphical representation of plans for identifying the type of cervical cancer. It demonstrates that proposed RBM is 10.61%, 9.98%, 14.61% and 16.27% more precise than RVDLNN, ELM, SVM and RF classifier respectively for 250 image dataset. The proposed RBM precision is increased on the basis of its high specificity and True positive rate in **Table 4.**

Table 4. Recall

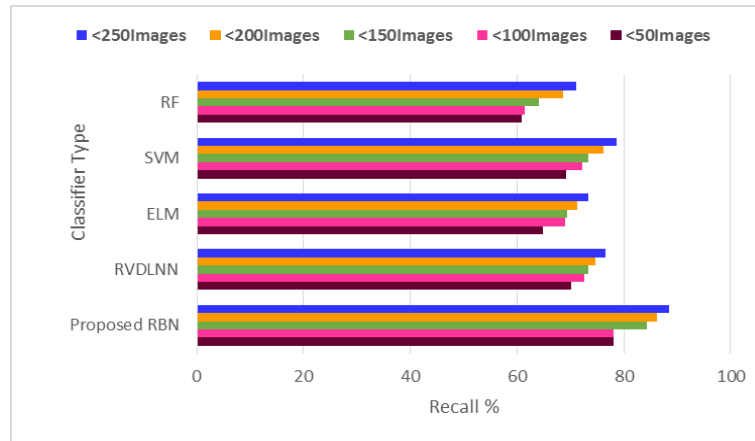| Image Set	Proposed RBN	RVDLNN	ELM	SVM	RF
50	78.14	70.15	64.76	69.15	60.88
100	78.14	72.54	69.01	72.14	61.46
150	84.34	73.33	69.45	73.41	64.15
200	86.15	74.74	71.36	76.11	68.67
250	88.56	76.48	73.44	78.61	71.11

Fig 7. Recall results of Classifiers

Fig 7 depicts the graphical representation of the recall execution of recognition of the type of cervical cancer plans. It demonstrates that RBN outperforms RVDLNN, ELM, SVM and RF classifier respectively for 250 image dataset by 12.08%, 15.12%, 9.95% and 17.45%, respectively. Due to its low error rate and high specificity, RBN is the subject of numerous investigations in **Table 5**.

Table 5. F-measure

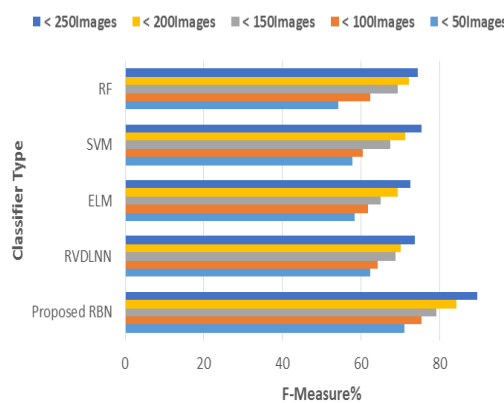| Image Set	Proposed RBN	RVDLNN	ELM	SVM	RF
50	71.11	62.33	58.4	57.85	54.31
100	75.36	64.21	61.78	60.57	62.36
150	79.19	68.89	65.12	67.54	69.34
200	84.25	70.11	69.44	71.19	72.13
250	89.7	73.8	72.51	75.4	74.58

Fig 8. F-measure results of Classifier

The F-measure execution of cervical malignant growth acknowledgment plans graphicaldepiction is showed up in **Fig 8**. It demonstrates that proposed RBMreview is 15.9 %, 17.19%, 14.3%, and 15.12%RVDLNN, ELM, SVM and RF classifier respectively for 250 image dataset. Exhibitions of the proposed scheme have improved clustering's high positive rate and more positive rate.

After training, the generalization performance of the classifiers is evaluated with the test data. The performance measures of the classifiers have been shown in **Table 1-5** along with the graph **Fig 4-8**. From, each iteration, increasing

the number of image dataset for testing to find the performance measures such as Testing accuracy, Specificity, Precision, Recall and F-measure. From the above **Table 1-5**, we analysed that our proposed system outperforms than the existing classifiers such as RVDLNN, ELM, SVM and RF.

Table 6. Overall performance numerical values for Pap Smear Cancer Identification

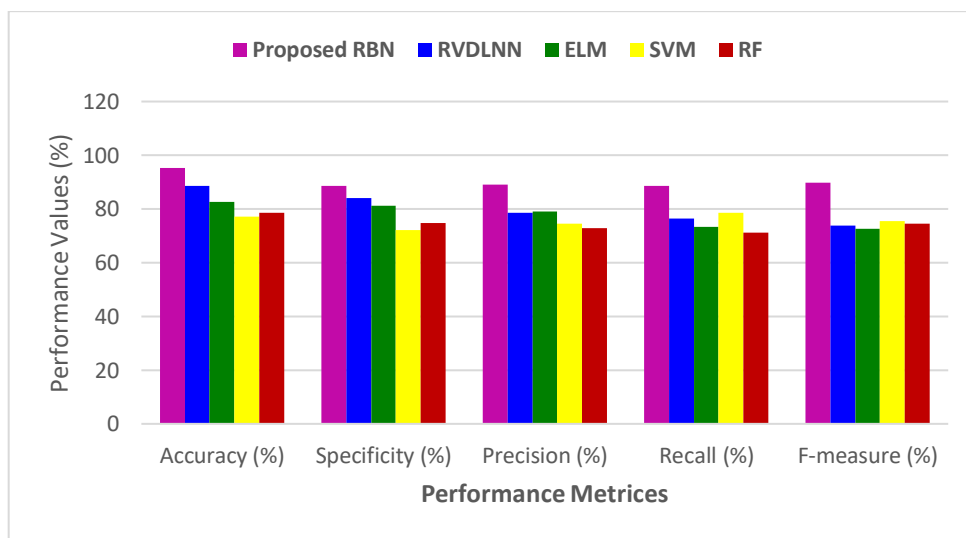| Performance Measures	Proposed RBN	RVDLNN	ELM	SVM	RF
Accuracy (%)	95.3	88.5	82.65	77.26	78.6
Specificity (%)	88.6	84.17	81.25	72.19	74.81
Precision (%)	89.13	78.52	79.15	74.52	72.86
Recall (%)	88.56	76.48	73.44	78.61	71.11
F-measure (%)	89.7	73.8	72.51	75.4	74.58

Fig 9. Prediction of overall performances for Pap Smear Cancer Classification schemes

The obtained results are significant when compared to all previous studies because the highest accuracy achieved with the same dataset is 95.3%. The RBM classifier was used to achieve this level of accuracy. This study used an existing structure and proposed a new structure to improve accuracy, sensitivity, and precision for all classes, despite the fact that the literature has focused on traditional methods. Additionally, the suggested approach is quick and precise. The proposed structure is straightforward, original, and precise, and the amount of time required to test a single new image does not exceed milliseconds, which is acceptable for use in medical applications**. Fig 9** shows Prediction of overall performances for Pap Smear Cancer Classification schemes. **Table 6** shows Overall performance numerical values for Pap Smear Cancer Identification.
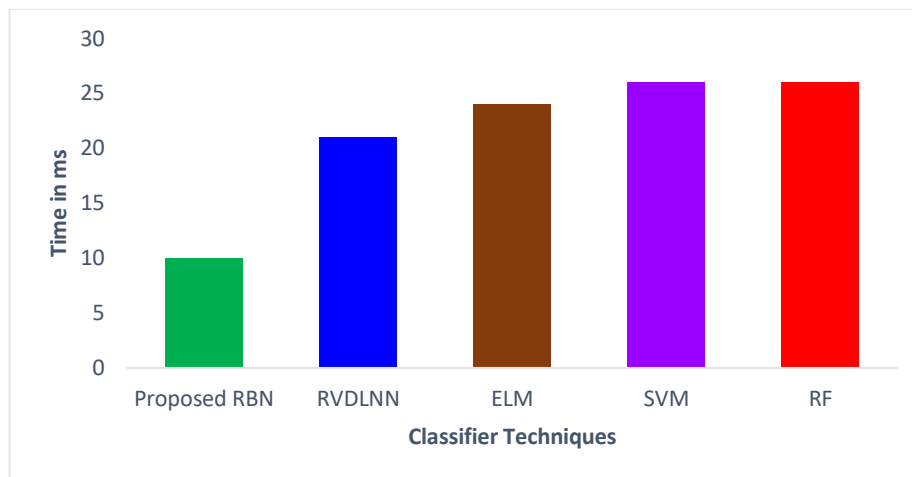
Fig 10. Computation Time

The computation time required by various algorithms, such as Proposed RBM, RVDLNN, ELM SVM, and RF, as well as the proposed work of RBM, is shown to be minimal in **Fig 10**. It is claimed in this paper that the proposed method is more accurate than other methods currently in use.

V. CONCLUSION

Based on Deep learning-based RBM classification, this study proposes Cervix Cancer Identification with the Pap Smear Test. Anisotropic Diffusion Filter with Histogram Equalization is used for reducing the commotion without eliminating the edges and by improving the contrast of image for better segmentation. An Improved Weighted FCM was utilized for determining the optimal cluster center of an image in the segmentation which leads to extract the features efficiently. The RBM classifier is utilized for classification. When compared to other classifiers like RVDLNN, ELM, SVM, and RF, the result of the proposed RBM achieved an accuracy of 95.3%. In the future, test the Pap smear image using a different algorithm that makes it easier and faster to find the cervical cancer earlier.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. World Health Organization. WHO Cancer Regional Profile 2020; International Agency for Research on Cancer: Lyon, France, 2020; pp. 1–2.

[2]. R. Huddart, "Faculty Opinions recommendation of Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.," Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature, Aug. 2021, doi: 10.3410/f.734004835.793587632.

[3]. J. Ferlay et al., "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," International Journal of Cancer, vol. 144, no. 8, pp. 1941–1953, Dec. 2018, doi: 10.1002/ijc.31937.

[4]. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. CA Cancer J. Clin. 2021, 71, 7–33.

[5]. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. CA Cancer J. Clin. 2020, 70, 7–30

[6]. W. Azani Mustafa, A. Halim, M. Wafi Nasrudin, And K. Shakir Ab Rahman, "Cervical cancer situation in Malaysia: A systematic literature review," BIOCELL, vol. 46, no. 2, pp. 367–381, 2022, doi: 10.32604/biocell.2022.016814.

[7]. N. B. Nahrawi, W. A. Mustafa, and S. N. A. Mohd Kanafiah, "Knowledge of Human Papillomavirus (HPV) and Cervical Cancer among Malaysia Residents: A Review," Sains Malaysiana, vol. 49, no. 7, pp. 1687–1695, Jul. 2020, doi: 10.17576/jsm-2020-4907-19.

[8]. Geetha and S. Suganya, "Enhancing the Classification of Pap Smear Images using ENN – TLBO classification Method," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 6, pp. 553–558, Apr. 2020, doi: 10.35940/ijitee.f3708.049620.

[9].   C. Zhang, D. Jia, Z. Li, and N. Wu, "Auxiliary classification of cervical cells based on multi-domain hybrid deep learning framework," Biomedical Signal Processing and Control, vol. 77, p. 103739, Aug. 2022, doi: 10.1016/j.bspc.2022.103739.

[10]. J. Li et al., "Cervical cell multi-classification algorithm using global context information and attention mechanism," Tissue and Cell, vol. 74, p. 101677, Feb. 2022, doi: 10.1016/j.tice.2021.101677.

[11]. G. Kumawat, S. K. Vishwakarma, P. Chakrabarti, P. Chittora, T. Chakrabarti, and J. C.-W. Lin, "Prognosis of Cervical Cancer Disease by Applying Machine Learning Techniques," Journal of Circuits, Systems and Computers, vol. 32, no. 01, Aug. 2022, doi: 10.1142/s0218126623500196.

[12]. N. Lavanya Devi and P. Thirumurugan, "Cervical Cancer Classification from Pap Smear Images Using Modified Fuzzy C Means, PCA, and KNN," IETE Journal of Research, vol. 68, no. 3, pp. 1591–1598, Nov. 2021, doi: 10.1080/03772063.2021.1997353.

[13]. Serra, J.Image (1982). "Analysis and Mathematical Morphology, Vol I. Academic Press, London

[14]. Gomila, C., Meyer, F. (1999). "Levelings in vector spaces". In Proceedings of the IEEE, Conference on Image Processing, Kobe, Japan. Vol. 2, pp.929-933

[15]. Chen, Q., Zhou, C., Luo, J., Ming, Dfast (2004). "Segmentation of high-resolution satellite approach". In Proceedings of the IWCIA, Auckland, New Zealand 621–630

[16]. Serra,J.Image (1982). "Analysis and Mathematical Morphology, Vol I. Academic Press, London.

[17]. Gomila, C., Meyer, F. (1999). "Levelings in vector spaces". In Proceedings of the IEEE, Conference on Image Processing, Kobe, Japan. Vol. 2, pp.929-933.

[18]. U. Khusanov and C. H. Lee, "Image enhancement based on local histogram specification," Journal of Korean Institute of Intelligent Systems, vol. 23, no. 1, pp. 18-23, 2013. http://dx.doi.org/10.5391/JKIIS.2013.23.1.18

[19]. Steven L Eddins, Richard E.woods, Rafael C "digital image processing" (2009)

[20]. M. Yasmin, S. Mohsin, M. Sharif, M. Raza, and S. Masood, "Brain image analysis: A survey," World Applied Sciences Journal, vol. 19, pp. 1484-1494, 2012.

[21]. M. Sharif, M. Y. Javed, and S. Mohsin, "Face recognition based on facial features," Research Journal of Applied Sciences, Engineering and Technology, vol. 4, pp. 2879- 2886, 2012.

[22]. Chen, Q., Zhou, C., Luo, J., Ming, Dfast, "Segmentation of high-resolution satellite approach". In Proceedings of the IWCIA, Auckland, New Zealand 621–630, 2004.

[23]. Nida N., Irtaza A., Javed A., Yousaf M.H., Mahmood M.T., "Melanoma Lesion Detection and Segmentation Using Deep Region Based Convolutional Neural Network and Fuzzy C-Means Clustering", Int. J. Med. Inform. 2019; 124:37–48. Doi: 10.1016/j.ijmedinf.2019.01.005.

[24]. Zhang, Wenyuan&Guo, Xijuan& Huang, Tianyu& Liu, Jiale& Chen, Jun., "Kernel-Based Robust Bias-Correction Fuzzy Weighted C-Ordered-Means Clustering Algorithm", Symmetry. 11. 753. 10.3390/sym11060753. 2009.

[25]. E. Agliari, F. E. Leonelli, and C. Marullo, "Storing, learning and retrieving biased patterns," Applied Mathematics and Computation, vol. 415, p. 126716, Feb. 2022, doi: 10.1016/j.amc.2021.126716.

[26]. M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images," 2018 25th IEEE International Conference on Image Processing (ICIP), Oct. 2018, doi: 10.1109/icip.2018.8451588.

[27]. Y. Marinakis, M. Marinaki, G. Dounias, J. Jantzen, and B. Bjerregaard, "Intelligent and nature inspired optimization methods in medicine: the Pap smear cell classification problem," Expert Systems, vol. 26, no. 5, pp. 433–457, Nov. 2009, doi: 10.1111/j.1468-0394.2009.00506.x.