



IMPROVED CERVICAL CANCER CLASSIFICATION BY WEIGHTED MAJORITY VOTING WITH SAMPLING

Lavanya Devi N^{*1}, C N Savithri², M Jayanthi³, S M Asha Banu⁴

¹Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, Dindigul, Tamil Nadu, India.

²Department of Electronics and Communication Engineering, Sri Sairam Engineering College, Chennai, Tamil Nadu, India.

³Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.

⁴Department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore, India.

¹<https://orcid.org/0000-0002-3908-0162>, ²<https://orcid.org/0000-0001-6291-229X>,

³<https://orcid.org/0000-0002-6510-6629>, ⁴<https://orcid.org/0000-0002-1250-6397>

Email: *lavanyadevinadarajan@gmail.com, savithri.ece@sairam.edu.in, jayanthisathish1012@gmail.com, ashabanu005@gmail.com

ARTICLE INFO

Article History

Received: January 9, 2026

Reviewed: February 3, 2026

Accepted: April 15, 2026

Published: May 29, 2026

Keywords:

Human Papilloma Virus,
Cervical cancer,
Weighted majority voting
classifier,
Principal Component Analysis,
Cross validation

ABSTRACT

Developing countries have a high rate of cervical cancer due to the detection of cancer at the stage of regional spread of the disease rather than detection at the early stage. Early detection of the disease by a periodical pap smear test is the most effective way of reducing the rate of cervical cancer. This research paper shares a novel model for the automated classification of cervical cancer with weighted majority voting. Upsampling and downsampling are done to overcome the inefficiency due to an imbalance in the database. Size, statistical, intensity, and texture features are extracted from the modified fuzzy c-means segmented image. Principal Component Analysis (PCA) is done on the extracted features to rank and identify the most significant features. Supported Vector Machine (SVM), K Nearest Neighbour (KNN), and Naïve Bayes (NB) are tuned to the kernel function and a number of cross validations and are given as the input to the weighted majority voting system. The performance of the proposed method with up sampling, down sampling, and no sampling are compared, and it was found that the weighted majority voting with upsampling has the highest classification accuracy of 99%.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The cervix is the lower bottom neck part of the uterus that joins to the vagina. Human Papillomavirus is the root source of cervical cancer [1] which is sexually transmitted. The risk factors for cervical cancer are smoking, miscarriage prevention drugs, weak immune system, sexual activity at an early stage, and multiple sexual partners. Most of the time the body's immune system will successfully fight against the virus and it will go by itself. But in some cases, it will develop into cancer and if it is not identified at an early stage, a benign cell will become a malignant cell. The cervical cancer at the benign stage usually produces no symptoms or signs [2]. In the advanced stage after menopause or between periods, bloody or watery discharge from the vagina with a foul odor, pelvic pain, etc can be seen. The contribution of Indian women towards worldwide cervical cancer is 25.41%. Cervical cancer is the second most frequently occurring cancer among women in the age of 15-44 [3] and is second among developing countries [4]. Routine pap smear checks will help in the early spotting of cervical cancer [5]. The colposcopic view of the cervical cell samples is taken either by punch biopsy or endocervical curettage and is used for the detection of cervical cancer. In punch biopsy, a sharp tool is used to take some samples of cervical cancer and in endocervical curettage, a thin brush is used to scrap the cervical cell sample. These samples will be analyzed and if the presence of cancerous cells is found, then depending upon the stage of the cancer the doctor will start the treatment. For instance, different treatment techniques like surgery to remove the cervix alone, radiation therapy, surgery to remove the cervix and uterus, immunotherapy, chemotherapy, targeted therapy, and supportive (palliative) care are provided to the patient depending upon the stage of cancer. The conventional method of detecting cervical cancer is prone to error and it is tiresome.

As a result, the accuracy of the manual detection is very poor and the inaccuracy in the results will increase the mortality rate due to the disease [6], [7]. To overcome the limitation of manual detection, automated detection of cervical cancer can be used [8]. The major steps involved in the automated detection of cervical cancer are preprocessing, segmentation [9], [10], feature extraction [11] and classification [12]. The major contributions in this paper are,

- To develop an efficient and novel algorithm for the classification of the cervical cell into cancerous and noncancerous cells using an ensemble of ML based classifiers.
- To compare the performance of various sampling methods to address the consequences of imbalance database.
- Identify the prime features among the twenty one different features comprising size, statistical, intensity and texture feature.
- To compare the proposed method with the state-of-the-art for the classification of the cervical cell into cancerous and non cancerous cells.
- To evaluate the performance of the weighted and non weighted voting classifier.

The paper is ordered as given below. Section 2 gives an outline of the literature review, section 3 is related to the proposed method, section 4 briefs about the result and its discussion, and section 5 explains the conclusion and future work.

II. LITERATURE REVIEW

A method for the prediction of cancer cells in the breast, lungs, oral, blood, stomach, and cervix by combining decision tree and clustering techniques is done in [13]. Preprocessing techniques are applied to the database and the K-means clustering splits the input image into cancerous and non-cancerous classes. The authors have developed a prediction system by framing the decision tree by assigning weights to various risk factors like age, living area, anemia, hazards due to occupation, family history of cancer, education, and habits. Paper [14] suggests a novel approach for the classification of arrhythmic beats using SVM. Discrete wavelet transform is used for the feature extraction and different machine learning techniques are used for the classification. The paper [15] proposed a model for pulmonary nodule detection using a modified V-net and Support Vector Machine (SVM) and various CNN methods are employed to improve the False Positive cases. The authors have employed the LIDC-IDRI data set for the conduction of the experiment. Finally, the authors have concluded that the SVM is the best method for the reduction of False Positive cases. The paper [16] projected a system for the automated detection of epileptic seizure with the help of EEG signals combined with various machine learning classifiers like SVM, Naive Bayes (NB) and K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN). The classification experiment obtained the best accuracy for ANN. In paper [5] an enhanced approach for the prostate classification by fine tuning the SVM and using ridge regression for the feature selection is being suggested. Optimization of hyperparameters of the ridge regression model is also done to reduce the complexity of the algorithm. The optimization has improved the accuracy by 49.54% on average.

In paper [1], the researchers have put forward an algorithm for the automated detection of breast cancer which incorporates three SVM based feature selection techniques. In [17] proposed a method for the diagnosis of cervical cells using SVM and two upgraded forms of SVM with recursive feature extraction and PCA. The authors have identified thirty two factors that are considered to be the causes of cervical cancer and applied them to four different types of inputs like Schiller, biopsy, cytology, and Hinselmann. Accuracy, sensitivity, specificity, positive predictive accuracy, and negative predictive accuracy are used for computing the efficiency of the proposed algorithm. SVM with PCA turned out to be the most efficient algorithm. In [18] suggested a method for cervical cancer staging by Dictionary Learning. The major contributions of this research paper are i) finding the complement of axial and sagittal MR images of the cervix ii) computing the self similarity of the Dictionary Learning iii) computing the features. A method for the early detection and prevention of cervical cancer by Boruta Analysis is done in [19]. In this paper cervical cancer influential factors are listed which affect the demographic, medical, and historical factors. The algorithm creates a copy of the dataset and builds the shadow of the features, a random forest classifier is used for training the dataset, and significant features are identified by comparing the reorganized/shuffled copy of the features with the original data. Boruta will identify each feature as rejected, confirmed, and tentative.

In [20], breast cancer classification by different ML techniques like SVM, KNN, DT, RF, and LR are discussed and promising results are reported by SVM. In [21], a voting classifier with SVM, KNN, Deep Neural Network Multilayer Perceptron, and Hard Voting Classifier Techniques are deployed for the liver disease identification and the hard voting classifier turned up with the best results with an accuracy of 0.78, specificity of 0.94, F-score of 0.87, precision of 0.8. An improved KNN [22] is suggested for the data set which involves a large dataset and features. According to the authors, the performance of the traditional KNN reduces when large data samples are given as input to the system. In [23], the authors suggested an idea for the categorizing cervical cells using pap images by data mining technique and made a comparative study of different classification methods like Naïve Bayes, SVM, and Random Forest Tree. 38 characteristics of cells that help in the classification of the cell into cancerous and noncancerous are discussed in the paper. The accuracy, recall, precision, and ROC area of the above three classification algorithms are computed and compared. Random forest is found to be the best classifier according to the analysis made. In [24], the proposed average perceptron classifier for breast cancer detection reduced the false negative and false positive cases and achieved an accuracy score of 0.984. The authors have highlighted the significance of the threshold in reducing the false negative and false positive cases.

III. MATERIALS AND METHOD

The proposed model is carried out in MATLAB on Herlev database with 565 single cell pap smear images (460 abnormal and 105 normal). Figure 1 shows the flow diagram of the proposed method.

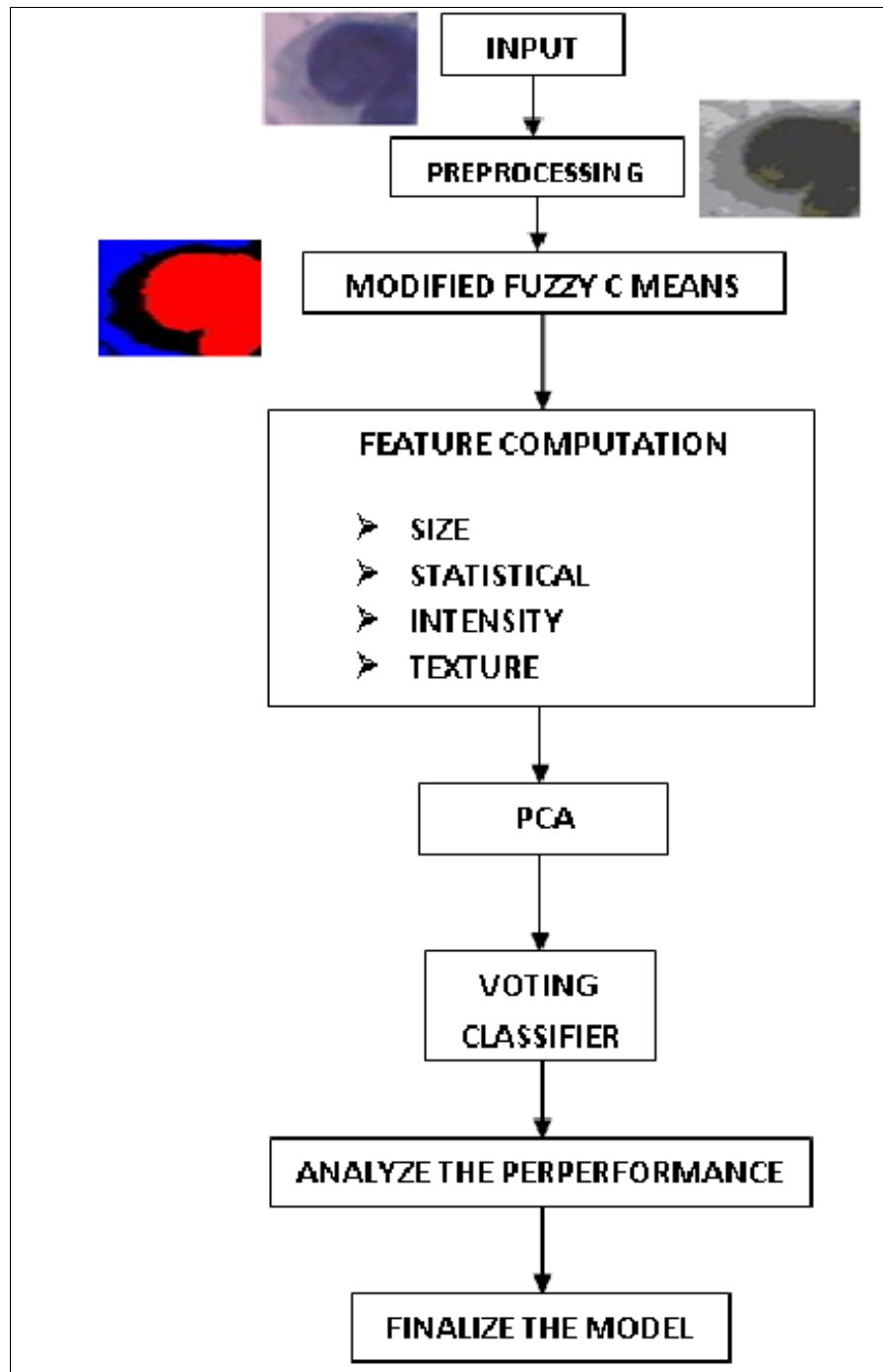


Figure 1: Proposed cervical cell classification algorithm.

Source: Authors, (2026).

III.1 SEGMENTATION BY MODIFIED FUZZY C- MEANS (MFCM)

Segmentation is the action of dividing the image into sub regions [25-28]. Modified Fuzzy C-means is used for partitioning the image into nucleus, cytoplasm, and background. In the conventional Fuzzy C-means algorithm, only the Euclidean distance between every pixel and the centroid is calculated. The pixel will be assigned to the class whose centroid is having minimum distance from the pixel. The conventional Fuzzy C-means calculates centroids for each partition (nucleus, cytoplasm, and background). The conventional Fuzzy C means does not consider the gray level value of the pixels in the image. The gray level value is an important parameter when grouping similar pixels as the dimension and orientation of each cell in the database are different. This method incorporates both the physical distance and also the gray level value in the algorithm. The modified fuzzy C-means algorithm sets a weight for Euclidean distance and the variation of the gray level value of each pixel from centroid and each pixel. The algorithm of the modified fuzzy C-means segmentation method is as follows:

Step 1: Initialize the membership function U .

Step 2: Initialize the number of clusters as 3 and compute the centroid using equation 1.

Step 3: Calculate the Euclidean distance between the centroid $c(i,j)$ and the pixel $x(i,j)$. Also, calculate the variation of the gray level value of each pixel with the gray level value of the centroid pixel. Apply weight for Euclidean distance as 0.25 and weight for gray level value difference as 0.75 by using equation 2.

Step 4: Compute the membership function using equation 3.

Step 5: Repeat steps 1-4 till the terminating condition calculated using equation 4 is reached. Otherwise, return to step 1.

Step 6: Apply pseudo coloring to obtain a better view of the segmented image.

Step 7: Apply the 'imfill' operation to fill the holes present in the image; apply 'imclearborder' operation to suppress the area lighter than the neighborhood.

$$c_j = \frac{\sum_{i=1}^N u_{ij}^2 x_i}{\sum_{i=1}^N u_{ij}^2} \quad (1)$$

$$d(i, j) = \sum_{i=1}^N \sum_{j=1}^N ((0.25(\sqrt{(i_c - i_x)^2 + (j_c - j_x)^2}) + 0.75(abs(c(i, j) - x(i, j)))) \quad (2)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

$$\|U^{(k+1)} - U^k\| < \varepsilon \quad (4)$$

III.2 FEATURE COMPUTATION

Features are parameters that help in identifying the correct class of the image [29]. Size, statistical, intensity, and texture features are used for classifying the dataset into normal and abnormal classes. The size feature as shown in table 1 details the size and shape of normal and abnormal cells as it may be different for both of them. Statistical features as shown in table 1 are related to the momentum and randomness in the image. Intensity features as shown in table 2 are related to the intensity values of the pixels and the texture features compute the irregularity, fineness, coarseness, etc. The texture features shown in table 2 are extracted by the Gray Level Co-occurrence Matrix (GLCM) calculated for [0 1] distance for an angle 00. GLCM computes the occurrence of a gray level intensity value x along with another gray level intensity value y in a particular spatial relationship.

Table 1: Size and statistical features for classification.

	Features	Significance of features
Size	Perimeter	Measures the circumference of a region
	Solidity	A portion of area of the region compared to the convex hull
	Extent	Quantity of pixels in the region to quantify pixels in the bounding box
	Equiv Diameter	Diameter of the circle that gives the same area as that of the region
	Convex Area	Quantity of pixels in the convex image
	Euler Number	Difference between quantity of objects in the region and quantity of holes in those objects
	Filled Area	Quantity of pixels present in the filled image
	Area	Quantity of pixels in a particular region
Statistics	Eccentricity	Eccentricity of the ellipse whose second order moment is equal to the region
	Orientation	Angle connecting the major axis of the ellipse and x axis whose second order moment is equal to that of the region
	Major Axis Length	Major axis length of the ellipse whose normalised value of the second order central moment is same
	Minor Axis Length	Minor axis length of the ellipse whose normalised value of the second order central moment is same
	Entropy	Measures the randomness of the image

Source: Authors, (2026).

Table 2: Intensity and texture features for classification.

Intensity	Acellular ratio	Quantity of pixels in the region with high intensity value	$Acellularratio = \frac{\text{number of pixels in the background}}{\text{total number of pixels}}$
	Cytoplasm ratio	Quantity of pixels in the region with medium intensity value	$Cytoplasm ratio = \frac{\text{number of pixels in cytoplasm}}{\text{total number of pixels}}$
	Nuclei ratio	Number of pixels in the region with high intensity value	$Nuclei ratio = \frac{\text{number of pixels in the nucleus}}{\text{total number of pixels}}$
	Ratio of Background to nuclei area	Number of pixels present in the background of image to pixels present in the nucleus	$\text{Ratio of background to nuclei area} = \frac{\text{number of pixels in background}}{\text{pixels in the nucleus}}$
Texture	Correlation	Computes the joint probability of existence of particular intensity pixel pairs	$Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j}$
	Contrast	Computes the variation in local change in GLCM	$Contrast = \sum_{i,j} i - j ^2 p(i, j)$
	Homogeneity	Computes the proximity of a particular element in the Gray Level Co-occurrence Matrix to the element in the diagonal	$Homogeneity = \sum_{i,j} \frac{p(i, j)}{1 + i - j }$
	Energy	Computes the uniformity	$Energy = \sum_{i,j} p(i, j)^2$

Source: Authors, (2026).

III.3 PCA AND EQUALIZATION OF DATA

Principal Component Analysis (PCA) of these features is computed by calculating the eigenvalues and eigenvectors. PCA ranks the features for classification and the eigenvalues obtained for the features are 2.5775, 0.0739, 0.0264, 0.0055, 0.0032, 0.0003, 0.0002, 0.0001 for Homogeneity, Energy, Correlation, Contrast, Perimeter, Extent, Solidity, EquivDiameter respectively (All other eigen values are negligibly low and are not considered). All four texture features are identified as the principal component and only those features are considered for classification. This process helps in reducing the computational time by reducing the number of features while retaining the maximum amount of information.

III.4 CLASSIFICATION

The classification algorithm categorizes the input image into normal and abnormal cells based on the features computed [30].

III.4.1 Majority Voting Method

The majority voting algorithm takes the vote from each algorithm while classifying each image into normal and abnormal [31], [32]. The voting algorithm classifies the image as normal if the majority of the algorithm classifies the image as normal and classifies the image as abnormal if the majority of the algorithm classifies the image as abnormal. The proposed majority voting algorithm considered SVM, Naïve Bayes, and KNN classifiers. Instead of directly feeding the output of SVM, Naïve Bayes, and KNN to the majority voting method, the classification algorithms are optimized for kernel function and the number of cross validations. For K fold cross validation [33] the whole dataset is divided into K divisions. Among the K divisions, one set is utilized for testing and the remaining K-1 set is used for training. In the next iteration, another set is chosen for testing and the remaining set is used for training. This process continues till all the sets are chosen for both testing and training. As the vast variation in the number of normal and abnormal images in the database affects the performance of the system [34], the following cases are analyzed for 5, 10, and 15 fold cross validation.

CASE A: 460 abnormal images and 105 normal images – Single cells of Herlev data set

CASE B: Up sampled database with 460 abnormal images and 460 normal images – 105 normal images in the database are randomly selected multiple times to make the number of normal images in the database 460 equalize the imbalance in the database.

CASE C: Down sampled database with 105 abnormal and 105 normal images - 105 abnormal images are randomly selected among the 460 abnormal images in the database to equalize the imbalance in the database.

Naïve Bayes (NB) classifiers are based on probability and Bayes theorem with the assumption of having independent features. Depending on the kernel used for classification, Gaussian Naïve Bayes or Kernel Naïve Bayes classifiers can be modeled and the accuracy of the Gaussian and Kernel Naïve Bayes for 5 fold, 10 fold and 15 fold are noted in table 3. The NB classifier works well for the dataset with high predictor samples [35], [36]. Gaussian Naïve Bayes classifier with 15 fold with an accuracy of 96.6% for no sampling, Kernel Naïve Bayes with 15 fold with an accuracy of 95% for upsampling and 96.7% for downsampling are selected for voting classification for case A, B and C respectively.

Support Vector Machine (SVM) generates N dimensional hyperplanes to classify the data. The value of N depends on the number of features selected. The important property of SVM classifiers is that they maximize the margin, minimize the classification error and give promising results for binary classification. As the performance of SVM is strongly affected by the selection of kernel function, different kernel functions like cubic, linear, quadratic, and coarse are tested for 5 fold, 10 fold, and 15 fold cross validation in this experiment and are shown in table 3. Cubic SVM with 15 fold marks the maximum accuracy for no sampling, upsampling, and down sampling and is selected for cases A, B, and C.

K Nearest Neighbor (KNN) [37] classifies the data based on their similarity. KNN algorithm which is also known as lazy learner does not learn much during the training phase; instead it merely stores the data and then it classifies the data based on the similarity. Performance of KNN does not depend on the type of data distribution. The performance of Weighted, Coarse, medium and Cosine KNN for 5 fold, 10 fold and 15 fold with no sampling, up sampling and down sampling are shown in table 3. Medium KNN with 15 fold is selected for case A and weighted KNN is selected for case B and C as these cases recorded the maximum accuracy.

Table 3: Accuracy for different classifiers with no, with and without sampling.

Sampling	Algorithm	5 Fold	10 Fold	15 Fold
No sampling	Gaussian Naïve Bayes	96.6	96.6	96.6
	Kernel Naïve Bayes	96.6	96.6	96.3
	Linear SVM	96.6	96.6	96.5
	Quadratic SVM	96.8	96.6	96.8
	Cubic SVM	96.8	96.6	96.8
	Coarse SVM	94.9	94.9	94.7
	Medium KNN	96.8	97.0	97.0
	Coarse KNN	93.8	94.0	94.0
	Cosine KNN	94.3	94.2	94.5
	Weighted KNN	96.1	95.9	95.9
Up sampling	Gaussian Naïve Bayes	94.7	94.7	94.6
	Kernel Naïve Bayes	95.0	95.0	95.0
	Linear SVM	95.3	95.2	95.1
	Quadratic SVM	95.8	95.7	96.3
	Cubic SVM	96.2	96.3	96.3
	Coarse SVM	94.6	94.2	94.8
	Medium KNN	96.3	96.4	96.4
	Coarse KNN	94.5	94.2	94.5
	Cosine KNN	95.2	94.8	94.7
	Weighted KNN	98	98.2	98.2
Down sampling	Gaussian Naïve Bayes	94.7	95.2	95.7
	Kernel Naïve Bayes	96.7	96.2	96.7
	Linear SVM	94.3	95.7	96.2
	Quadratic SVM	94.7	95.2	95.2
	Cubic SVM	95.2	95.2	96.2
	Coarse SVM	94.7	94.8	94.8
	Medium KNN	95.2	95.7	96.7
	Coarse KNN	95.7	95.2	94.8
	Cosine KNN	96.2	95.7	95.7
	Weighted KNN	95.7	96.2	97.1

Source: Authors, (2026).

IV RESULT AND DISCUSSION

The majority voting method applied on cases A, B, and C with the fine tuned Naïve Bayes, SVM, and KNN are applied for the following

- i) No weight for the algorithm
- ii) With weight for the algorithm- According to the performance ranking of the algorithm, the weight [38] for KNN, SVM, and Naïve Bayes is fixed as 0.5, 0.25, and 0.25 respectively.

Accuracy, precision, sensitivity, specificity, F1-score [39] and Miss Classification Rate are computed by equations 5 to 10 respectively. Equations 5 to 10 are calculated by using True Negative (TN), False Negative (FN), True Positive (TP), and False Predictive (FP). TN value shows a non-cancerous cell predicted as non-cancerous, the TP value shows a cancerous cell predicted as cancerous, the FP value shows a non-cancerous cell predicted as a cancerous cell, the FN value shows a cancerous cell predicted as a non-cancerous cell.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

$$F1 - Score = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \quad (9)$$

$$Miss \text{ Classification Rate}(MCR) = \frac{FP + FN}{TP + TN + FP + FN} \quad (10)$$

Table 4: Performance of the majority voting classifier with and without weight.

Sampling	With/Without weight	Accuracy	TP	FN	FP	TN	Precision	Sensitivity	Specificity	F1 score	MCR
no sampling	With weight	98.05	101	5	6	453	94.39	95.28	98.69	94.83	0.019
	Without weight	97.87	100	6	6	453	94.34	94.34	98.69	94.29	0.0212
Up sampling	With weight	99	460	0	9	451	98.08	100	98.04	99.03	0.0098
	Without weight	98.8	458	2	9	451	98.07	99.57	98.04	98.81	0.0195
Down Sampling	With weight	98.09	103	2	2	103	98.09	98.09	98.09	98.09	0.0190
	Without weight	97.62	102	3	2	103	98.08	97.14	98.09	97.61	0.0238

Source: Authors, (2026).

The table 4 shows the performance comparison of no sampling, upsampling, and downsampling with and without weight. Weighted majority voting classifier with upsampling shows better performance (accuracy- 99%, precision – 98.08%, sensitivity – 100%, specificity – 98.04%, F1-score – 99.03%, MCR – 0.0098) compared to others as the classification system learns equally about the normal and abnormal images during the training phase. The majority voting classifier with down sampling shows better performance compared to the no sampling case as both the classes (normal and abnormal) are treated equally during the training phase. Table 5 shows the proficiency of the proposed algorithm concerning the accuracy, sensitivity, and specificity compared to the state-of-the-art.

Table 5: Comparison of accuracy, sensitivity and specificity of the proposed algorithm with the state-of-the-art.

	Algorithm used	Algorithm applied on	Parameter		
			Accuracy	Sensitivity	Specificity
Proposed method	Weighted voting with tuned parameters and up sampling	Herlev data base	99	100	98.04
Wu et al. [17]	SVM	Hinselmann	93.97	100	89.96
		Schiller	90.18	98.7	84.83
		Biopsy	94.13	100	90.21
		Cytology	92.75	100	87.92
	SVM with recursive feature extraction	Hinselmann	93.69	100	89.49
		Schiller	90.18	98.73	84.63
		Biopsy	94.03	100	90.05
		Cytology	92.37	100	87.28
	SVM with PCA	Hinselmann	93.79	100	89.65
		Schiller	90.18	98.99	84.30
		Biopsy	94.03	100	90.05
		Cytology	92.46	100	87.44
Arya et al. [40]	Fuzzy	Generated data base	93.3	-	-
	Fuzzy with PCA		95	-	-
Abdoh et al. [41]	Random Forest Classifier With SMOTE and Feature Reduction Techniques	Biopsy test	96.06	94.55	97.51
Rayavarapu et al. [42]	Voting	Biopsy test	99	100	96
	DNN		95	98	38

Source: Authors, (2026).

V. CONCLUSION

We proposed a weighted majority voting method for the classification of the cervical cells into cancerous and noncancerous. ML based algorithms like SVM, NB, and KNN are tuned with respect to kernel functions and a different number of cross validations and are given as the input to the weighted majority voting classifier. Fixing of weight of each classifier was based on the individual performance of SVM, KNN, and NB. To address the problem of imbalance in the data, up-sampling and down-sampling of data were done, and noted that there was an improvement in the performance of the model when the sampling technique was adopted. The main limitation of the proposed machine learning algorithm is its dependency on the features selected and the model is not suitable for overlapping cells in the database. In feature, some advanced techniques for feature selection can be employed. Also, to overcome the dependency of the classification algorithm on features, the classification of cervical cancer can be done by a deep learning approach. Deep learning algorithms may improve the accuracy of classification and may work efficiently on all the images in the database.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Methodology: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Investigation: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Discussion of results: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Writing – Original Draft: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Writing – Review and Editing: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Resources: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

Supervision: Lavanya Devi N, C N Savithri, M Jayanthi, S M Asha Banu.

VII. ACKNOWLEDGMENTS

The authors are very much thankful to the department and institution for the valuable support while carrying out this work.

VIII. REFERENCES

- [1] Lavanya Devi, N., & Thirumurugan, P. Cervical cancer classification from pap smear images using modified fuzzy C means, PCA, and KNN. IETE Journal of Research 2022; 68(3), 1591-1598.
- [2] Ramachandran, P., Girija, N., & Bhuvaneswari, T. (2014). Early detection and prevention of cancer using data mining techniques. International Journal of Computer Applications, 97(13), 48-53.

- [3] Sreedevi, A., Javed, R., & Dinesh, A. Epidemiology of cervical cancer with special focus on India. *International journal of women's health* 2015; 405-414.
- [4] Nirmal, R., Yun, C., Le, M., Paripoonnanonda, P., & Yi, J. Digital health game on cervical health and its effect on american women's cervical cancer knowledge. In *2013 IEEE International Games Innovation Conference (IGIC) 2013*; 191-198 IEEE.
- [5] Varan, M., Azimjonov, J., & MaÇal, B. Enhancing prostate cancer classification by leveraging key radiomics features and using the fine-tuned linear svm algorithm. *IEEE Access* 2023.
- [6] Kuriakose, R., Aggarwal, A., Sohi, R. K., Goel, R., Rashmi, N. C., & Gambhir, R. S. Patient safety in primary and outpatient health care. *Journal of family medicine and primary care* 2020;9(1), 7-11.
- [7] Cao, W., Chen, H. D., Yu, Y. W., Li, N., & Chen, W. Q. (2021). Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics *Chinese medical journal* 2020; 134(07), 783-791.
- [8] Verma, M. Artificial intelligence role in modern science: aims, merits, risks and its applications. *Artificial Intelligence* 2023; 7(5).
- [9] Liu, Y., Zhang, P., Song, Q., Li, A., Zhang, P., & Gui, Z. Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. *IEEE Access* 2018; 6, 53709-53721.
- [10] Tareef, A., Song, Y., Huang, H., Feng, D., Chen, M., Wang, Y et al. Multi-pass fast watershed for accurate segmentation of overlapping cervical cells. *IEEE transactions on medical imaging* 2018; 37(9), 2044-2059.
- [11] Abdoh, S. F., Rizka, M. A., & Maghraby, F. A. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. *IEEE Access* 2018; 6, 59475-59485.
- [12] Abda, O., & Naimi, H. (2025). Enhanced Brain Tumor MRI Classification Using Stationary Wavelet Transform, ResNet50V2, and LSTM Networks. *ITEGAM-JETIA*, 11(51), 127-133.
- [13] Ramachandran, P., Girija, N., & Bhuvaneswari, T. Early detection and prevention of cancer using data mining techniques. *International Journal of Computer Applications* 2014; 97(13).
- [14] Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., & Kumar, R. ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. *IEEE Access* 2018; 6, 9767-9773.
- [15] Ye, Y., Tian, M., Liu, Q., & Tai, H. M. Pulmonary nodule detection using v-net and high-level descriptor based svm classifier, *Ieee Access* 2020; 8, 176033-176041.
- [16] Mardini, W., Yassein, M. M. B., Al-Rawashdeh, R., Aljawarneh, S., Khamayseh, Y., & Meqdadi, O. Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers. *IEEE Access* 2020; 8, 24046-24055.
- [17] Wu, W., & Zhou, H. Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access* 2017; 5, 25189-25195.
- [18] Bnoui, N., Rekik, I., Rhim, M. S., & Amara, N. E. B. Cross-view self-similarity using shared dictionary learning for cervical cancer staging. *IEEE Access* 2019; 7, 30079-30088.
- [19] Gupta, P., Jindal, I., & Goyal, A. Early Detection and Prevention of Cervical Cancer. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT) 2019*; (pp. 1-4). IEEE.
- [20] Naji, M. A., El Filali, S., Aarika, K., Benlahmar, E. H., Abdelouahid, R. A., & Debauche, O. Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, 2021; 191, 487-492.
- [21] Anthonysamy, V., & Babu, S. K. Multi Perceptron Neural Network and Voting Classifier for Liver Disease Dataset. *IEEE Access* 2023.
- [22] Xing, W., & Bei, Y. Medical health big data classification based on KNN classification algorithm. *Ieee Access* 2019; 8, 28808-28819.
- [23] Kumiawati, Y. E., Permanasari, A. E., & Fauziati, S. Comparative study on data mining classification methods for cervical cancer prediction using pap smear results. In *2016 1st International Conference on Biomedical Engineering (IBIOMED) 2016*; (pp. 1-5). IEEE.
- [24] Birchha, V., & Nigam, B. Performance analysis of averaged perceptron machine learning classifier for breast cancer detection. *Procedia Computer Science* 2023; 218, 2181-2190.
- [25] Meng, F., Guo, L., Wu, Q., & Li, H. A new deep segmentation quality assessment network for refining bounding box based segmentation. *IEEE Access* 2019; 7, 59514-59523.
- [26] Song, J., Xiao, L., & Lian, Z. Contour-seed pairs learning-based framework for simultaneously detecting and segmenting various overlapping cells/nuclei in microscopy images. *IEEE Transactions on Image Processing* 2018; 27(12), 5759-5774.
- [27] Lu, C., Chelikani, S., Jaffray, D. A., Milosevic, M. F., Staib, L. H., & Duncan, J. S. Simultaneous nonrigid registration, segmentation, and tumor detection in MRI guided cervical cancer radiation therapy. *IEEE transactions on medical imaging* 2012; 31(6), 1213-1227.
- [28] V. Rajinikanth, S. Kadry, and Y. Nam, "Convolutional-neural-network assisted segmentation and svm classification of brain tumor in clinical mri slices," *Information Technology and Control*, vol. 50, no. 2, pp. 342–356, 2021, doi: 10.5755/j01.itc.50.2.28087.
- [29] Liu, J., Peng, Y., & Zhang, Y. A fuzzy reasoning model for cervical intraepithelial neoplasia classification using temporal grayscale change and textures of cervical images during acetic acid tests. *Ieee Access* 2019;7, 13536-13545.
- [30] Li, C., Chen, H., Zhang, L., Xu, N., Xue, D., Hu, Z et al. Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning. *Ieee Acces* 2019; 7, 90378-90397. doi: <https://doi.org/10.1109/ACCESS.2019.2924467>

- [31] Jamil, H., Qayyum, F., Iqbal, N., Jamil, F., & Kim, D. H. Optimal ensemble scheme for human activity recognition and floor detection based on AutoML and weighted soft voting using smartphone sensors. *IEEE Sensors Journal* 2022; 23(3), 2878-2890.
- [32] Batool, A., & Byun, Y. C. Towards Improving Breast Cancer Classification using an Adaptive Voting Ensemble Learning Algorithm. *IEEE Access* 2024.
- [33] Naseem, U., Rashid, J., Ali, L., Kim, J., Haq, Q. E. U., Awan, M. J., & Imran, M. An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers. *Ieee Access* 2022; 10, 78242-78252.
- [34] Blanchard, A. E., Gao, S., Yoon, H. J., Christian, J. B., Durbin, E. B., Wu, X. C et al. A keyword-enhanced approach to handle class imbalance in clinical text classification. *IEEE journal of biomedical and health informatics* 2022; 26(6), 2796-2803.
- [35] Chaurasia, V., Pal, S., & Tiwari, B. B. Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology* 2018; 12(2), 119-126.
- [36] Radak, M., Lafta, H. Y., & Fallahi, H. Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. *Journal of Cancer Research and Clinical Oncology*, 2023; 149(12), 10473-10491.
- [37] Arboleda, E., & Rodrin, S. (2025). Classification of Prominent Cacao Pod Diseases Using Multi-Feature Visual Analysis And K-Nearest Neighbors Algorithm. *Itegamgn1 Sistemas e Publicacoes Ltd.*, 11(51), 28-34.
- [38] Martínez-Licort, R., León, C. D. L. C., Agarwal, D., Sahelices, B., De La Torre, I., Miramontes-González, J. P et al. Breast Carcinoma Prediction through Integration of Machine Learning Models 2024; *IEEE Access*.
- [39] Devi, N. L., & Thirumurugan, P. A literature survey of automated detection of cervical cancer cell in Pap smear images. *World Review of Science, Technology and Sustainable Development* 2022; 18(1), 74-82.
- [40] Arya, M., Mittal, N., & Singh, G. Fuzzy-based classification for cervical dysplasia using smear images. In *Recent Findings in Intelligent Computing Techniques* 2018; (pp. 441-449). Springer, Singapore.
- [41] Abdoh, S. F., Rizka, M. A., & Maghraby, F. A. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. *IEEE Access* 2018; 6, 59475-59485.
- [42] Rayavarapu, K., & Krishna, K. K. Prediction of Cervical Cancer using Voting and DNN Classifiers. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* 2018; 1-5, IEEE.